FDA Briefing Document
Psychopharmacologic Drugs Advisory Committee (PDAC) Meeting

March 29, 2016

The attached package contains background information prepared by the Food and Drug Administration (FDA) for the panel members of the advisory committee. The FDA background package often contains assessments and/or conclusions and recommendations written by individual FDA reviewers. Such conclusions and recommendations do not necessarily represent the final position of the individual reviewers, nor do they necessarily represent the final position of the Review Division or Office. We bring NDA 207318, pimavanserin immediate-release, film-coated oral tablets for the treatment of psychosis associated with Parkinson's disease, to this Advisory Committee in order to gain the Committee's insights and opinions. The background package may not include all issues relevant to the final regulatory recommendation and instead is intended to focus on issues identified by the Agency for discussion by the advisory committee. The FDA will not issue a final determination on the issues at hand until input from the advisory committee process has been considered and all reviews have been finalized. The final determination may be affected by issues not discussed at the advisory committee meeting.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 DIVISION MEMORANDUM

M E M O R A N D U M

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

DATE: 2 March 2016

FROM: Mitchell V. Mathis, M.D.
Director
Division of Psychiatry Products, HFD-130

TO: Members of the Psychopharmacologic Drugs Advisory Committee (PDAC)

SUBJECT: March 29, 2016 Meeting of the PDAC

This one-day PDAC meeting will focus on the use of pimavanserin for the treatment of psychosis in Parkinson's disease (PDP). The application for pimavanserin (ACP-103), trade name Nuplazid, is sponsored by Acadia Pharmaceuticals. Pimavanserin is an inverse agonist at serotonin receptor subtype 5-$HT_{2A,}$ and unlike other antipsychotics, it has a notable lack of interaction with dopamine receptors. The applicant believes that this unique pharmacologic profile allows for the treatment of psychosis without worsening the motor symptoms of Parkinson's disease.

No drugs are currently approved for the treatment of PDP. Acadia applied for Breakthrough Therapy Designation and it was granted, and a Priority Review is currently underway for this product. PDP is characterized by multiple motor and non-motor symptoms, and psychosis can occur in as many as 40 percent of patients with Parkinson's disease at some time during the course of their illness. PDP has been associated with nursing home placement, which itself has been associated with increased morbidity and mortality for these patients; therefore, a drug to prevent psychosis in this population without worsening motor function would, indeed, represent an important treatment advance.

The applicant has submitted a single positive Phase 3 pivotal trial (ACP-103-020) that evaluated the efficacy, tolerability, and safety of pimavanserin in patients with PDP. In addition, the results of three other randomized controlled trials were submitted as supportive information, but none of these three was statistically positive on its primary endpoint. Although the Division usually requires evidence of efficacy from more than one positive, adequate, and well-controlled trial, it is within our authority to rely on one robustly positive trial, especially when we have supportive evidence from the early part of the development program.

Although Trial 020 is strongly statistically positive, there has been a fair amount of consideration among the review team members about how to characterize the clinical meaningfulness of the statistically significant change seen in this trial. Clinical meaningfulness is not always easy to assess in psychiatric drug trials and that is the case with this development program—statistical significance is undoubtedly a component of clinical meaningfulness, but not all of it. Clinical meaningfulness is difficult to quantify, but it is, in the most general sense, an effect produced by the drug that matters to the patient or to the treating clinician. It is worth noting that the mean drug effect does not fully describe the benefit. There will be patients with a response above the mean, and how these larger responses compare between drug and placebo is of interest. While the applicant will present evidence that the treatment effect represents a clinically meaningful change, the Division's medical officer will present his interpretation of the same data and reach a different conclusion.

Interpretation of the clinical meaningfulness of a drug effect has to be considered in light of its toxicity. In other words, if a drug produces a statistically significant effect on a reasonable endpoint in a trial, and if that drug is completely without safety signals, then we need not spend too much time defining clinical meaningfulness because benefit, however small, outweighs zero risk, and even a small benefit in a disabling disease is valuable. If, as is the case with pimavanserin, the safety risk is substantial (increased serious morbidity and increased mortality), the evaluation of clinical meaningfulness is critical to making a risk-benefit decision. It will be important for the Division to have the discussion about clinical meaningfulness as it relates to the pimavanserin development program.

The applicant will present its case to the PDAC, then FDA will present the reviewer's interpretation of the data, and then we will ask the members of the PDAC to answer three key questions to understand the benefit and risks of pimavanserin for the treatment of PDP.

Draft Points to Consider

1. Has the applicant provided substantial evidence of effectiveness for pimavanserin for the treatment of psychosis associated with Parkinson's disease?
2. Has the applicant adequately characterized the safety profile of pimavanserin?
3. Do the benefits of pimavanserin for the treatment of psychosis associated with Parkinson's disease outweigh the risks of treatment?

2 REGULATORY BACKGROUND

Pimavanserin was developed under IND 68,384 for the treatment of psychosis associated with Parkinson's disease (PDP). The clinical program consisted of four randomized, controlled trials for safety and efficacy, which utilized the Scale for the Assessment of Positive Symptoms (SAPS) as the primary efficacy variable. This scale, which was originally developed for schizophrenia, failed to show a statistically significant improvement in psychosis symptoms in three of the trials. Considering these results, Acadia met with the Agency in April, 2010, to discuss their clinical program and modifications of the design for a subsequent pivotal trial. Of note, the SAPS was adjusted to a 9-item scale to improve clinical relevance for PDP. The resulting trial (ACP-103-020) was statistically positive.

Following the completion of this trial, Acadia again met with the Agency to gain agreement that an NDA would be accepted for filing on the basis of data from a single, strongly positive trial with supportive safety and efficacy data from their earlier trials. FDA usually requires evidence of more than one positive, adequate, and well-controlled trial for drug approval; however, the Agency's 1998 Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products describes a number of circumstances under which a single trial may serve as the basis of approval for a new drug. This development program also received Breakthrough Therapy Designation in August, 2014.

Additional meetings between the Agency and Acadia occurred throughout development to discuss key aspects of the clinical, nonclinical, and CMC programs. Relevant regulatory history and milestones are highlighted below for reference:

- [2006 June & September respectively] **Type C and End-of-Phase 2 Clinical/ Nonclinical meetings held**. Phase 2 efficacy data from Trial ACP-103-006 were discussed and early agreement was reached on the endpoints, measures, and trial design for the Phase 2b/3 Studies, ACP-103-012 and ACP-103-014.
- [2010 April] **Written correspondence issued in lieu of Type C Meeting**. The modified trial design and primary endpoint were agreed upon for Trial ACP-103-020.
- [2013 April] **Type C meeting held**. The Division agreed to file an NDA on the basis of strongly positive data from ACP-103-020 with supportive data from previous trials.
- [2014 June] **Pre-NDA meeting held**. The organization of the NDA, its review path, and specific aspects of the content and presentation of clinical and nonclinical data were agreed upon, including the SAPS and general structure of the Integrated Summary of Effectiveness (ISE).
- [2014 August] **Breakthrough Therapy Designation granted**.

3 EFFICACY AND SAFETY EVALUATION FOR PIMAVANSERIN

3.1. Review of Pivotal Efficacy Trial ACP-103-020

The applicant completed four randomized controlled trials of pimavanserin in PDP. Trial ACP-103-020 is the only statistically positive controlled trial in the applicant's development program for PDP.

Trial ACP-130-020 employed a primary efficacy variable that was derived from the SAPS and is referred to as the SAPS-PD (Scale of Positive Symptoms-Parkinson's Disease). This scale was designed based on a factor analysis of the failed trials. The questions on the SAPS that showed the most favorable change in the failed trials were compiled into a 9-item scale that measured the domains of hallucinations and delusions. This scale was applied to trial ACP-130-020 prospectively and scored by a central rating system.

Overview and Objective

Trial ACP-103-020, "A Multi-Center, Placebo-Controlled, Double-Blind Trial to Examine the Safety and Efficacy of Pimavanserin in the Treatment of Psychosis in Parkinson's Disease," was designed to assess the efficacy and safety of pimavanserin 34 mg daily in the treatment of PDP as measured by a decrease in the severity and/or frequency of hallucinations and/or delusions.

Trial Design

ACP-103-020 was a six-week, multi-center, randomized, double-blind, placebo-controlled trial.
Pimavanserin (ACP-103) was administered at 34 mg and compared to a placebo arm.
The design called for approximately 100 patients per treatment arm.
The trial was conducted on an outpatient basis with visits performed as follows:

> Screening Visit 1, Day 1 (Baseline), Day 15, Day 29 and Day 43 with a follow-up visit (Day 71) 4 weeks after the last regular trial visit for those patients who do not continue into an open-label extension protocol.

At the screening visit, a trained member of the site staff met with the patient's caregiver to devise a structured plan of social interaction for the patient and caregiver to follow at home. This brief non-pharmacologic psychosocial counseling was intended to help the patient and caregiver to manage the symptoms and provide standard of care prior to the blinded investigational treatment phase. Following the screening visit, patients were to receive two follow-up phone calls (at about 3 and 7 days from the screening visit) to review the plan and evaluate progress.

Only those patients who met entry criteria at baseline were to be randomized to receive 34 mg pimavanserin or matching placebo for the 6-week treatment period.

The trial population was to include approximately 200 patients who were to meet the following criteria:

Key Inclusion Criteria:

1. Male or female of 40 years of age or older with a clinical diagnosis of idiopathic Parkinson's disease with a minimum duration of 1 year, defined as the presence of at least three of the following cardinal features, in the absence of alternative explanations or atypical features: rest tremor, rigidity, bradykinesia and/or akinesia, postural and gait abnormalities.
2. Female patients must have been of non-childbearing potential (defined as either surgically sterilized or at least 1 year post-menopausal) or must have agreed to use a clinically acceptable method of contraception (such as intrauterine device [IUD], diaphragm, oral, injectable [e.g. Depo-Provera] or implantable contraception [e.g. Norplant System]), for at least one month prior to randomization, during the trial, and one month following completion of the trial.
3. Patients must have had psychotic symptoms that developed after the diagnosis of Parkinson's disease was established. These symptoms must have included visual hallucinations and/or auditory hallucinations, and/or delusions.
4. Psychotic symptoms were to have been present for at least one month and the patient must have been actively experiencing psychotic symptoms each week during the month prior to the Screening visit.
5. Symptoms severe enough to warrant treatment with an antipsychotic agent; documented at screening by items A and B of the Neuropsychiatric Inventory (NPI), and defined as a score of 4 or greater on either the Hallucinations (Frequency x Severity) or Delusions (Frequency x Severity) scales OR a total combined score of 6 or greater.
6. At the baseline visit, patient must have had a SAPS Hallucinations or Delusions global item (H7 or D13) score $\geq$3 AND a score >3 on at least one other non-global item using the modified 9-item SAPS Hallucinations and Delusions domains.
7. Patient must have had a clear sensorium at trial entry (i.e., oriented to time, person, and place).
8. Patient must have been on stable dose of anti-Parkinson's medication for 1 month prior to Day 1 (Baseline) and during the trial.

Key Exclusion Criteria:

1. Patient with psychotic symptoms (hallucinations and delusions) which could be better explained as a part of a toxic, metabolic or infection-induced delirium/encephalopathy, psychosis due to substance abuse, psychosis associated with schizophrenia, bipolar disorder or psychotic depression.
2. Patient with a history of significant psychotic disorders prior to or concomitantly with the diagnosis of Parkinson's disease including, but not limited to, schizophrenia or bipolar disorder.

Trial Endpoints

The primary endpoint was the SAPS-PD, a 9-item scale consisting of a subset of items from the 20-item SAPS (Andreasen, 1984). Separate items are rated from 0 (absent) to 5 (severe), for a total possible score on the SAPS-PD ranging from 0 to 45.

The SAPS was designed to measure positive symptoms in schizophrenia. Positive symptoms include delusions, hallucinations, abnormalities in language and behavior, and disordered thought processes. Two of the SAPS subscales, hallucinations and delusions, were to be administered in this trial. The entire 20-item assessment was to be administered at Day 1 (Baseline), Day 15, Day 29 and Day 43. If patients terminated before Day 43 the scale was to be administered at the early termination visit.

For trial inclusion and analysis purposes, nine of the 20 Hallucinations (H) and Delusions (D) items were to be used. These items are:

- H1 Auditory Hallucinations
- H3 Voices Conversing
- H4 Somatic or Tactile Hallucinations
- H6 Visual Hallucinations
- H7 Global Rating of Severity of Hallucinations

- D1 Persecutory Delusions
- D2 Delusions of Jealousy
- D7 Ideas and Delusions of Reference
- D13 Global Rating of Severity of Delusions

The selection of these domains and items was based principally on their relevance to the specific symptomatology of the PDP population and their utility, as demonstrated in a *post hoc* analysis of the previously failed studies of pimavanserin. (Please refer to review by the Clinical Outcomes Assessment [COA] Staff.)

A centralized rater service was employed to conduct the SAPS assessments. They were used to control for inter-rater variability across sites and to obtain a “blinded” rating of patient symptom severity and change.

Secondary Efficacy: The Clinical Global Impression-Severity (CGI-S) is a clinician-rated scale that measures the patient’s current illness state and overall clinical state on a 1 (normal, not at all ill) to 7-point (extremely ill) scale.

Secondary Efficacy: The Clinical Global Impression-Improvement (CGI-I) is a clinician-rated scale that measures the patient’s change from the initiation (baseline) of treatment on a 1 (very much improved) to 7-point (very much worse) scale.

Secondary Safety and Function (Motor Control): The Unified Parkinson’s Disease Rating Scale (UPDRS) II+III is a clinical rating scale that measures the patient’s current Parkinson’s disease

state. The score was derived as the sum of the 27 items from activities of daily living and motor examination, with a range of 0 to 108.

The secondary measures of efficacy, the CGI-S and CGI-I, were assessed by trial investigators blinded to the SAPS-PD results. The primary endpoint was change from baseline in SAPS-PD total score at the end of Week 6. The change from baseline for pimavanserin was compared to placebo.

Patient Disposition

Overall, 199 patients were randomly assigned to treatment, including 94 patients in the placebo group and 105 in the pimavanserin 34 mg group. Of these, 87 (92.6%) of patients in the placebo group and 89 (84.8%) in the pimavanserin 34 mg group completed 6 weeks (42 days) of double-blind treatment.

- Across the treatment groups, 11.6 % patients discontinued the trial, with twice the rate of discontinuation in the pimavanserin 34 mg daily group. The most common reasons for discontinuation were Adverse Events (AEs) in 2 (2.1%) patients in the placebo group and 10 (9.5%) patients in the pimavanserin 34 mg group.
- A similar percentage of patients in the placebo and pimavanserin 34 mg groups discontinued the trial due to voluntary withdrawal of consent (2.1% vs 2.9%, respectively).

Table 1. Trial ACP-103-020 Patient Demographics Intent to Treat Analysis

Demographic Parameters	Placebo (N=90)	Pimavanserin 34 mg daily (N=95)
Sex (n, %)		
Male	52 (58%)	64 (67%)
Female	38 (42%)	31 (33%)
Age (years)		
Mean (SD)	72.4 (7.9)	72.4 (6.6)
Median	72.0	72.0
Min, max	53, 90	56, 85
Age Group (n, %)		
≥ 40 - < 65 years	11 (12%)	11 (12%)
> 65 - < 75 years	50 (56%)	53 (56%)
≥ 75 years	29 (32%)	31 (33%)
Race (n, %)		
White	85 (94%)	90 (95%)
Black or African American	1 (1%)	1 (1%)
Asian	0	0
Hispanic	2 (2%)	4 (4%)
Other	2 (2%)	0

Source: Applicant Table S20-1.2.2 and page 84/2137 Table 5-4 ISE NDA 207-318

Table 2. Selected Screening and Baseline Characteristics: ITT Analysis

Selected Screening and Baseline Characteristics	Placebo (N=90)	Pimavanserin 34 mg (N=95)
Neuropsychiatric Inventory (NPI), sum of delusion and hallucination domain scores, range 2 to 24		
Delusions (frequency X severity), range 1-12		
Mean (SEM)	4.9 (0.4)	4.8 (0.4)
SD	4.1	4.2
Median (min, max)	6.0 (0, 12)	6.0 (0, 12)
Hallucinations (frequency X severity), range 1-12		
Mean (SEM)	7.3 (0.3)	7.1 (0.3)
SD	2.8	2.8
Median (min, max)	8.0 (0, 12)	8.0 (0, 12)
NPI-H+D Score[b]		
Mean (SEM)	12.2 (0.6)	11.8 (0.6)
SD	5.3	5.9
Median (min, max)	12.0 (4, 24)	10.0 (4, 24)
Mini-Mental Status Exam (MMSE) Range = 0 to 30, higher = better		
Mean (SEM)	26.6 (0.3)	26.0 (0.3)
SD	2.4	2.6
Median (min, max)	27.0 (21, 30)	26.0 (21, 30)
Categorical, n (%)[d]		
<25	21 (23%)	29 (31%)
≥25	69 (77%)	66 (70%)
SAPS-PD (modified 9-item Scale for the Assessment of Positive Symptoms [SAPS] hallucinations and delusions score, range=0 to 45)		
Mean (SEM)	14.7 (0.6)	15.9 (0.6)
SD	5.5	6.1
Median (min, max)	14.0 (6, 30)	15.0 (6, 33)
SAPS-H+D (SAPS 20-item hallucinations and delusions score, range=0 to 100)		
Mean (SEM)	15.8 (0.7)	17.5 (0.8)
SD	6.5	7.6
Median (min, max)	14.0 (6, 37)	16.0 (6, 38)
GSAPS-H+D (Combined SAPS hallucinations and delusions global rating of severity score, range=0 to 10)		
Mean (SEM)	6.2 (0.2)	6.3 (0.2)
SD	1.9	2.0
Median (min, max)	6.0 (3, 10)	6.0 (3, 10)
Unified Parkinson's Disease Rating Scale (UPDRS) Parts II+III, range 0 to 160		

Selected Screening and Baseline Characteristics	Placebo (N=90)	Pimavanserin 34 mg (N=95)
Mean (SEM)	52.6 (1.8)	51.5 (1.8)
SD	17.1	17.6
Median (min, max)	51.5 (10.5, 100.0)	48.8 (21.5, 104.0)
Clinical Global Impression Score (CGI-S) (range 0 to 7)		
Mean (SEM)	4.3 (0.1)	4.3 (0.1)
SD	0.9	0.9
Median (min, max)	4.0 (2, 6)	4.0 (1, 6)
Time Since First PDP Symptom (months)		
Mean (SEM)	36.4 (4.2)	30.9 (3.1)
SD	39.6	30.0
Median (min, max)	24.4 (3, 292)	18.4 (2, 168)
Time Since PD Diagnosis (months)		
Mean (SEM)	127.5 (8.4)	115.6 (8.1)
SD	19.9	78.6
Median (min, max)	110.1 (20,412)	99.6 (14, 376)

Data source: From NDA 207-318 Submission Tables 14.1.2.2.1, 14.1.2.3.1, 14.1.2.4.1, 14.1.2.5.1, and 14.2.2.2.1.
Abbreviations: GSAPS-H+D=Combined SAPS hallucinations and delusions global rating of severity, score range=0 to 10;

[a] Score was derived as (frequency x severity) and was evaluated only if the symptom was present; score range was 1 to 12 for each domain.

Efficacy Results – Primary Endpoint

Trial ACP-103-020 demonstrated that pimavanserin is statistically superior to placebo at decreasing symptoms of psychosis associated with Parkinson's disease as measured by the SAPS-PD. Table 3 displays the results of the statistical analyses from the SAPS-PD (the primary efficacy variable) as well as secondary and exploratory efficacy variables.

Table 3. Summary of Efficacy at Day 43: All Scales, Domains, or Other Item Clusters

	Measure	Rater	Population	Analysis[a]	LSM Treatment Δ[b]	95% Confidence Intervals	p-value
ANTI-PSYCHOTIC EFFICACY							
Primary	SAPS-PD	Independent (Central)	ITT	MMRM	-3.1	(-4.9, -1.2)	0.001
			PP	MMRM	-3.2	(-5.1, -1.3)	0.001
			ITT	LOCF	-2.9	(-4.8, -1.1)	0.002
			ITT	WOCF	-2.8	(-4.6, -0.9)	0.003
			All rand	WOCF/BOCF	-2.4	(-4.1, -0.6)	0.008
Supportive	SAPS-PD % Change	Independent (Central)	ITT	MMRM	-23.1%	(-36%, -10%)	<0.001
	SAPS H+D	Independent (Central)	ITT	MMRM	-3.4	(-5.4, -1.4)	0.001
	SAPS H+D % Change	Independent (Central)	ITT	MMRM	-23.5%	(-37%, -10%)	<0.001
	GSAPS-H+D Score	Independent (Central)	ITT	MMRM	-0.9	(-1.7, -0.2)	0.012
	SAPS H	Independent (Central)	ITT	MMRM	-2.1	(-3.5, -0.7)	0.003
	SAPS D	Independent (Central)	ITT	MMRM	-1.2	(-2.2, -0.1)	0.033
Secondary	CGI-I	Site Investigator	ITT	MMRM	-0.7	(-1.1, -0.3)	0.001
	CGI-I responder	Site Investigator	ITT	Chi-square test	23.3%	(9.3%, 37.2%)	0.002
	CGI-S	Site Investigator	ITT	MMRM	-0.6	(-0.9, -0.3)	<0.001
OTHER EFFICACY							
Exploratory	SCOPA–Night	Site Investigator	ITT	MMRM	-0.9	(-1.8, -0.02)	0.045
	SCOPA-Day wake	Site Investigator	ITT	MMRM	-1.2	(-2.2, -0.3)	0.012
	Caregiver Burden	Caregiver	ITT	MMRM	-4.3	(-7.0, -1.7)	0.002
	Caregiver Burden – categorical	Caregiver	ITT	CMH	N/A	N/A	0.004

Notes: SAPS-PD=sum of 9-item PD-adapted SAPS; SAPS-H+D=sum of 20-items for H and D domains, SAPS-H=sum of 7 items for H domain, SAPS-D=sum of 13 items for D domain, GSAPS-H+D=sum of the global item for each of the H and D domains (2 items total); MMRM=mixed model repeated measures analysis; OC=observed cases; ANCOVA=analysis of covariance; LOCF=last-observation-carried-forward; WOCF=worst-observation-carried-forward; BOCF=baseline-observation-carried-forward; CMH=Cochran-Mantel-Haenszel test; LSM=least square means

[a] MMRM refers to MMRM(OC) analyses; ANCOVA was used for all LOCF, WOCF and BOCF imputation methods, b LSM treatment Δ = pimavanserin minus placebo

From the table above, the SAPS-PD change was -23.1%. A 22-34% change has been defined as "minimally improved" (Leucht, et al., 2006). Therefore, as interpreted from these data, pimavanserin produced a minimally improved clinical change in patients with Parkinson's psychosis.

3.2. Review of Safety

3.2.1. Review of the Safety Database

Overall Exposure in the Pimavanserin Development Program at NDA Submission

The integrated safety database for pimavanserin includes 1592 patients from 18 trials (in addition to the ongoing open-label extension trial ACP-103-015). Across all enrolled patients, 1096 have been exposed to pimavanserin alone or in combination with adjunctive therapy; of these, 625 had Parkinson's disease/PDP (616 with PDP), 177 had schizophrenia, and 294 were healthy volunteers. Total subject exposure in PDP is approximately 825 person-years (the majority at the dose of 34 mg) and the longest single exposure exceeds 8 years.

Among 498 patients with PDP who have been enrolled in open-label safety extension studies, 338 have received once daily pimavanserin for >6 months, 278 have exceeded 12 months of treatment and 141 have exceeded 24 months of treatment.

Across all studies, the majority of patients received pimavanserin doses from 8.5 to 34 mg: 764 patients were exposed to pimavanserin 34 mg, 343 patients to pimavanserin 17 mg, and 140 patients to pimavanserin 8.5 mg (4 additional patients received pimavanserin 25 mg). Ten patients were exposed to pimavanserin 42.5 mg, 54 patients to pimavanserin 57 mg, 72 patients to pimavanserin 68 mg, and 40 patients to pimavanserin 85 mg. Eight or fewer patients received doses from 102 mg to 255 mg. A total of 698 patients received placebo or placebo/adjunctive therapy.

Table 4. Trial Settings of Exposure to Pimavanserin in the Safety Population

Clinical Trial Groups	Pimavanserin (n=1096)	Active Control (n= 269)	Placebo-only (n=210)
Normal Volunteers	294	0	146
Controlled trials conducted for PDP	412	0	64 (only placebo, no extension trial)
All other PDP exposures than controlled trials	213	0	0
Controlled trials conducted for Schizophrenia	177	269	0 (add-on studies)

Source page 84/12167 ISS NDA 207318 Table 4-2

Table 5. Cumulative Long-term Patient Exposure to Pimavanserin

Number of patients exposed to the Pimavanserin: 1096		
>=6 months	>=12 months	>=24 months
N=338	N=278	N=141

Source: Applicant Table PDPLT 1-6 and page 85/12167 Table 4-5 NDA 207318

3.2.2. Relevant characteristics of the safety population

The total number of exposures in the pimavanserin development program is below the ICH guideline for total exposures; however, the exposure database is within the ICH recommendation for the number of patients exposed for periods of 6 months and one year. Two-hundred and two patients with PDP were exposed to the 34 mg daily dose in the 6-week controlled trial population (PDP6). This sample of patients compared to their appropriate control group demonstrates more than double the risk of death and serious adverse events (SAE) in the PDP6 trial population (Observed Risk of death or SAE is 2.38 times greater [95% CI 1.00 to 5.73, p=0.05]) for 34 mg pimavanserin vs. placebo. The demographic characteristics for this population are outlined in the tables below.

The population of patients enrolled in the 6-week clinical trials (Placebo-controlled 6-week Studies Population [PDP6]: ACP-103-012, ACP-103-014, and ACP-103-020) was generally representative of patients with PDP (Tables 6 and 7). The adverse event tables for labeling would most appropriately be calculated from this population for the following reasons:

1. This was the time period that was required to demonstrate efficacy; therefore, it is the minimum point at which a short-term comparison of benefit versus risk can be made.
2. The population of patients with PDP is frail and death and serious adverse events are reasonably expected to increase with greater amounts of time. Time periods for comparisons must be equal when comparing drug to placebo.
3. 6 weeks is the longest placebo controlled period of exposure in the pimavanserin exposure database in the PDP population.
4. It is the largest sample of PDP patients available in the development program to compare.

Table 6. Demographic Characteristics of PDP6 Population-Age, Sex, BMI

	Placebo (N=231)	PIM 8.5 mg (N=140)	PIM 17 mg (N=41)	PIM 34 mg (N=202)	All PIM (N=383)	Total (N=614)
Age (years)						
n	231	140	41	202	383	614
Mean (SD)	71.5 (8.8)	69.6 (8.4)	72.1 (8.2)	71.1 (7.3)	70.7 (7.8)	71.0 (8.2)
Median Min, Max	72.0 43, 90	70.0 44, 90	73.0 53, 88	71.0 40, 85	71.0 40, 90	71.0 40, 90
Age Category (years), n (%)						
40-64	45 (19.5)	37 (26.4)	7 (17.1)	35 (17.3)	79 (20.6)	124 (20.2)
65-75	105 (45.5)	69 (49.3)	16 (39.0)	108 (53.5)	193 (50.4)	298 (48.5)
>75	81 (35.1)	34 (24.3)	18 (43.9)	59 (29.2)	111 (29.0)	192 (31.3)
Age Group (years), n (%)						
≤75	150 (64.9)	106 (75.7)	23 (56.1)	143 (70.8)	272 (71.0)	422 (68.7)
>75	81 (35.1)	34 (24.3)	18 (43.9)	59 (29.2)	111 (29.0)	192 (31.3)
Sex, n (%)						
Male	134 (58.0)	89 (63.6)	24 (58.5)	144 (71.3)	257 (67.1)	391 (63.7)
Female	97 (42.0)	51 (36.4)	17 (41.5)	58 (28.7)	126 (32.9)	223 (36.3)
Weight (kg)						
n	229	137	41	202	380	609
Mean (SD)	73.6 (16.84)	71.7 (16.70)	71.4 (12.35)	75.3 (15.57)	73.6 (15.76)	73.6 (16.16)
Median	74.0	71.1	69.0	74.4	72.7	73.0
Min, Max	32, 150	41, 115	45, 97	44, 127	41, 127	32, 150
BMI (kg/m^2)						
n	229	136	41	200	377	606
Mean (SD)	26.2 (4.95)	25.5 (4.99)	26.7 (3.82)	26.0 (4.59)	25.9 (4.66)	26.0 (4.77)
Median Min, Max	25.9 15, 52	24.8 16, 42	26.6 18, 38	25.4 17, 43	25.3 16, 43	25.7 15, 52

Source: Applicant Table PDP6 1-2, NDA 207318, Integrated Summary of Safety, page 118/12167

Table 7. Demographics: Race, Ethnicity, Race Group, Area, and Geographic area PDP6 Population

	Placebo (N=231)	PIM 8.5 mg (N=140)	PIM 17 mg (N=41)	PIM 34 mg (N=202)	All PIM (N=383)	Total (N=614)
Race, n (%)						
White	209 (90.5)	124 (88.6)	41 (100.0)	183 (90.6)	348 (90.9)	557 (90.7)
Black	3 (1.3)	2 (1.4)	0	2 (1.0)	4 (1.0)	7 (1.1)
Asian	12 (5.2)	10 (7.1)	0	11 (5.4)	21 (5.5)	33 (5.4)
Other	7 (3.0)	4 (2.9)	0	6 (3.0)	10 (2.6)	17 (2.8)
Ethnicity, n (%)						
Hispanic	5 (2.2)	3 (2.1)	0	6 (3.0)	9 (2.3)	14 (2.3)
Non-Hispanic	226 (97.8)	137 (97.9)	41 (100.0)	196 (97.0)	374 (97.7)	600 (97.7)
Area, n (%)						
North America	156 (67.5)	62 (44.3)	18 (43.9)	149 (73.9)	229 (59.8)	385 (62.7)
Europe	65 (28.1)	68 (48.6)	23 (56.1)	43 (21.3)	134 (35.0)	199 (32.4)
India	10 (4.3)	10 (7.1)	0	10 (5.0)	20 (5.2)	30 (4.9)
Geographic area, n (%)						
North America	156 (67.5)	62 (44.3)	18 (43.9)	149 (73.8)	229 (59.8)	385 (62.7)
Outside North America	75 (32.5)	78 (55.7)	23 (56.1)	53 (26.2)	154 (40.2)	229 (37.3)

Source: Applicant Table PDP6 1-2, NDA 207318, Integrated Summary of Safety, page 120/12167

3.3. Safety Results

Deaths

The applicant states the following about the deaths that occurred during the pimavanserin development program (Source: ISS 9.3.1.1 All Treated Patients [Safety Analysis Population]-Introductory Statement):

> In total and across all studies, there were 57 deaths among the 1575 patients in the Safety Analysis Population (Table All 2-4.1) all occurring in PDP patients; 49 of the deaths occurred on treatment (i.e., within 30 days of last dose) and 8 deaths occurred more than 30 days after completion of dosing. Five deaths occurred during the double blind placebo controlled studies. Overall and among the deaths on treatment, a greater proportion occurred in pimavanserin-treated patients (48/901, 5.3%) compared to those who received placebo (1/210, 0.5%)…

Later in the NDA submission (ISS section 9.3.2.1.2) the applicant gives the number of 51/459 (11.1%) deaths among the PDP long term exposure patients. Though this number (51) includes patients who were more than 30 days post treatment, the denominator of 459 provides the most appropriate context for this application as the deaths all occurred in PDP patients.

The presence of psychotic symptoms increases the risk and expectation of mortality; however, evidence that hallucinations or other psychosis constitute an independent risk factor for mortality is presently lacking. A higher mortality was found in PD patients with hallucinations who had

entered nursing homes than in controls living in the community (Goetz & Stebbins, 1995). Psychosis is associated with dementia, which predicts increased mortality risk in PD (Levy G, et al., 2002).

In this analysis of death in pimavanserin trials, it is useful to examine the comparative rates of death and serious adverse events only in the placebo controlled trials that have comparable times of exposure for the purpose of exploring the comparative risk of death and serious adverse events associated with drug treatment. Comparing the five deaths (4 drug, 1 placebo) in the three randomized controlled trials, the estimated odds ratio is 2.94 (95% CI 0.28 to 148, p=0.61). If the one death on drug that occurred more than 60 days after initiation is excluded, the relative risk remains elevated at 2.39 (95% CI 0.18 to 128, p=0.81).

In this analysis of death in pimavanserin trials, it is useful to examine the comparative rates of death and serious adverse events only in the placebo controlled trials that have comparable times of exposure for the purpose of exploring the comparative risk of death and serious adverse events associated with drug treatment. Comparing the five deaths (4 drug, 1 placebo) in the three randomized controlled trials, the estimated odds ratio is 2.94 (95% CI 0.28 to 148, p=0.61). If the one death on drug that occurred more than 60 days after initiation is excluded, the relative risk remains elevated at 2.39 (95% CI 0.18 to 128, p=0.81).

The deaths that occurred in the pimavanserin development program do not appear to be pathologically unique relative to what is expected in the disease course of patients with PDP; however, they were numerically more frequent in the pimavanserin group versus the placebo group over the six-week treatment period. Because the numbers of patients in the studies are relatively small, this numerical difference could be attributed simply to chance; however, if this is merely a chance occurrence, then when one examines serious adverse events (including deaths) no trend or pattern in serious adverse events should be associated with this numerical difference. That said, when examining serious adverse events, a regression to an odds ratio of 1 does not occur as would be expected if this disproportionate number of deaths were a chance observation. On the contrary, there are a disproportionate number of serious adverse events in the PDP6 placebo controlled treatment population that reaches a level of statistical as well as clinical significance.

The following table lists the deaths in the PDP6 patient population. Deaths are neither unexpected nor obviously related to the study drug, given that these types of deaths occur routinely in this patient population.

Table 8. Deaths in Pimavanserin Placebo-Controlled Trials

Age/Sex	Dose (mg)	Days from Last Dose to Death	Days on Drug	Verbatim Term	Preferred Term
85 Male	Placebo	9	27	Cardio pulmonary Arrest	Cardio-Respiratory Arrest
61 Male	Pimavanserin 10mg	unknown	46	Probable Myocardial Infarction	Myocardial Infarction
76 Male	Pimavanserin 34 mg	1	9	Septic Shock	Septic Shock
74 Male	Pimavanserin 34 mg	7	38	Septicemia	Sepsis
84 Female	Pimavanserin 34 mg	32	29	Respiratory Distress	Respiratory Distress

Source: Applicant Table from Integrated Summary of Safety, page 9191

Likewise, the deaths that occurred during open-label pimavanserin exposure do not appear to have a unique or unifying underlying pathophysiological mechanism that would lead to the conclusion that a drug-related event was causing death. The fact that roughly 11% (51/459 in the pimavanserin long-term trial population [PDPLT]-Source ISS section 9.3.2.1.2) of the pimavanserin-exposed patient population with PDP died is likewise not unexpected given the context of death rates observed in the literature as noted above.

In summary, death associated with PDP is unfortunately a relatively common event. Mean survival times for patients with PDP vary in the literature from trial to trial, but reports of 2-4 years of survival are accepted in the literature as valid estimates. The deaths that occurred in the pimavanserin development program do not appear to be pathologically unique relative to deaths expected with the disease course of patients with PDP; however, they were numerically more frequent in the pimavanserin treatment group in the controlled trials, over the six-week treatment period.

Serious Adverse Events

A serious adverse event (SAE) is defined as an event resulting in death, life-threatening states, hospitalization (initial or prolonged), disability or permanent damage, or congenital anomaly or birth defects. SAEs may include other serious (important medical) events that do not fit the other listed outcomes, but the event may jeopardize the patient and may require medical or surgical intervention to prevent one of the other outcomes. Examples of such events include allergic bronchospasm requiring emergency treatment, serious blood dyscrasias or seizures/convulsions that do not result in hospitalization. As with death, SAEs are relatively common in the PDP patient population, which is generally elderly and frail. Aspiration, pneumonia, respiratory crisis, serious cardiovascular disease, sepsis, falls and their sequelae are common serious adverse events that occur in the PDP population as part of the course of the disease.

One death in the open-label trial population (PDPLT) was attributed to rhabdomyolysis; another patient had rhabdomyolysis reported as a serious adverse event and recovered. Rhabdomyolysis is a rare event and when it occurs in the context of new drug development it is usually attributed to the new drug treatment; however, "malignant syndrome," which includes rhabdomyolysis, is a well-documented condition in Parkinson's disease that is associated with a wide variety of drugs used in the treatment of Parkinson's disease as well as with physical stressors of the disease such as dehydration or constipation (Ikebe et al., 2003; Mizuno et al., 2003; Ogawa et al., 2012). Therefore, these two reports of rhabdomyolysis cannot readily be attributed to treatment with pimavanserin.

The observed risk (OR) in the controlled trial population stratified by trial, for serious adverse events (SAEs) in the pimavanserin development program is:

- 1.99 (95% CI 0.87 to 4.53, p=0.10) for all drug vs. placebo
- 2.38 (95% CI 1.00 to 5.73, p=0.05) for 34 mg vs. placebo
- 1.44 (95% CI 0.54 to 3.81, p=0.46) for less than 34 mg vs. placebo

The comparison of the pimavanserin 34 mg groups and the placebo groups in the PDP6 population is the most appropriate comparison to make in evaluating adverse events. The two groups are treated for the same amount of time, the numbers of adverse events increase with time, and pimavanserin 34 mg is the only dose that has proven efficacy.

Serious adverse events occurred in 16/202 (7.9%) patients taking pimavanserin 34 mg versus 8/231 (3.5%) placebo-treated patients in the PDP6 population (Table 9).

Table 9. Overall Adverse Event Summary for PDP Placebo-controlled 6-Week Studies (Population PDP6: ACP-103-012, ACP-103-014, and ACP-103-020)

	Double-blind Treatment					Open-label Treatment	Total (N=798) n (%)
	Placebo (N=231) n (%)	PIM 8.5 mg (N=140) n (%)	PIM 17 mg (N=41) n (%)	PIM 34 mg DB (N=202) n (%)	All PIM (N=383) n (%)	PIM 34 mg OL[b] (N=184) n (%)	
Any AE[a]	141 (61.0)	79 (56.4)	21 (51.2)	124 (61.4)	224 (58.5)	110 (59.8)	475 (59.5)
Any Severe AE	11 (4.8)	8 (5.7)	3 (7.3)	20 (9.9)	31 (8.1)	18 (9.8)	60 (7.5)
Any Serious AE	8 (3.5)	8 (5.7)	1 (2.4)	16 (7.9)	25 (6.5)	12 (6.5)	45 (5.6)
Any AE Leading to Discontinuation or Trial Termination	10 (4.3)	9 (6.4)	3 (7.3)	16 (7.9)	28 (7.3)	16 (8.7)	54 (6.8)
Any AE Resulting in Death	1 (0.4)	1 (0.7)	0	3 (1.5)	4 (1.0)	1 (0.5)	6 (0.8)

Source: Applicant, Table PDP6 2-1 and Page 155 of ISS

[a] Only adverse events (AEs) that occurred on or after the administration of first trial drug dose and before or on the last dose date (+30 days) are included.

[b] Includes adverse events only up to Day 72 for patients in ACP-103-015 that were in the placebo treatment group in the core studies ACP-103-012, -014, and -020.

Dropouts and/or Discontinuations Due to Adverse Effects

There was roughly twice the dropout rate in the pimavanserin 34 mg daily group compared to placebo in the PDP6 population: 10/231 (4.3%) patients dropped out of the placebo group due to an adverse event (AE) versus 16/202 (7.9%) in the pimavanserin 34 mg daily group.

Psychiatric disorders was the system organ class (SOC) with the highest incidence of AEs leading to discontinuation for both all pimavanserin (All PIM) and placebo groups (3.7% All PIM vs. 2.6% placebo), followed by nervous system disorders (1.8% All PIM vs. 0.4% placebo). AEs in all other SOCs occurred in ≤2 patients per arm. Within the psychiatric SOC, the most common AEs leading to discontinuation (>2 patients) in the double-blind pimavanserin 34 mg group were hallucination (4 patients [2.0%] vs. 1 patient [0.4%] placebo) and psychotic disorder (3 patients [1.5%] vs. 2 patients [0.9%] placebo).

Significant Adverse Events

Approximately twice as many patients in the All PIM group (8.1%) compared to the placebo group (4.8%) experienced AEs deemed “severe” in intensity by the investigator during the PDP placebo-controlled 6-week studies. The incidence of severe AEs increased with increasing pimavanserin dose: 5.7% for pimavanserin 8.5 mg, 7.3% for pimavanserin 17 mg, and 9.9% for pimavanserin 34 mg. In addition, 9.8% of patients experienced severe AEs in the first 6 weeks of open-label treatment with pimavanserin 34 mg after having received placebo in a blinded trial. As with the disproportionate increase of serious adverse events in the pimavanserin 34 mg daily group compared to the placebo group, there is no obvious unifying pathophysiologic process or unique adverse event that drives or dominates this disproportion.

Events Related to Pimavanserin Pharmacology or Clinical Experience

Cardiac QT interval prolongation: Based on clinical and preclinical experience, pimavanserin has the potential to increase QT Interval; this was explored in a thorough QT trial that was reviewed by the QT-Interdisciplinary Review Team (QT-IRT). Thorough QT trials are the standard for evaluating the risk of drug-related QT prolongation. The QT-IRT review is included in the background package.

A dose of 17 mg showed no effect on QT interval, whereas a dose of 68 mg produced an increase in QTcI that ranged from 10-14 msec. Of note, the 34 mg once daily therapeutic dose was not directly studied in the thorough QT study. Based on the linear pharmacokinetics (PK) of pimavanserin, however, the 68 mg dose is expected to provide a 2-fold margin over the therapeutic exposure. The CYP3A4/5 inhibitor ketoconazole increases the maximum plasma concentration (Cmax) of pimavanserin by 50% and triples area under the curve (AUC) in the single dose trial. The

effect of hepatic impairment and renal impairment on pimavanserin PK is unknown. Based on the concentration-QTc relationship, marginal QTc prolongation is expected at the therapeutic dose.

Weight-Loss Related AEs: The incidence of weight-loss and related events was similar for the All PIM group (1.8%) compared to the placebo group (1.7%). Despite the expectation of increased appetite and weight gain seen in other populations with drugs that possess 5-HT_{2C} inverse agonism, there were no weight gain-related events in the PDP6 population and weight loss was more prominent as an AE in the PDP studies. This may be because pimavanserin's potency at 5-HT_{2C} receptors is too low to mediate such effects and/or because cachexia and weight loss occur frequently in late-stage PD. In the PDP6 population, the frequency of reports for AEs in the category of weight-loss related events—decreased appetite, weight decreased and abnormal loss of weight—was numerically less in the pimavanserin 34 mg daily group than for placebo.

Neuroleptic Malignant Syndrome (NMS) Related Events: Rhabdomyolysis is a NMS-related event and was experienced by 3 patients (0.6%) in the PDPLT Population. As discussed above, it is difficult to attribute these cases of rhabdomyolysis to drug.

Suicidal Ideation and Behavior: The safety and efficacy studies in the pimavanserin clinical program were initiated prior to the release of the draft FDA guidance entitled, "Suicidal Ideation and Behavior: Prospective Assessment of Occurrence in Clinical Trials." For this reason, specific scales currently recommended to evaluate suicidality risk were not evaluated in trials of pimavanserin. The safety database was searched for any AE reports potentially related to suicidal ideation and behavior; the following events were found:

- In the PDP placebo-controlled 6-week studies (PDP6 Population), one patient in the pimavanserin 34 mg group experienced an AE of accidental overdose (medication unknown).
- In the open-label long-term studies, one patient made a suicide attempt during the PDP open-label long-term studies and two patients experienced an AE of suicidal ideation.

Adverse Events and Adverse Reactions

AEs experienced by ≥2% of patients (in the all pimavanserin treated patients [All PIM] or placebo groups) in the 6-week placebo-controlled PDP studies (Population PDP6) and partial (6-week) data from the open-label Trial 015 (for patients who received placebo in a core trial) are presented by MedDRA system organ class (SOC) and preferred term in Table 10.

The only pimavanserin dose with demonstrated efficacy in the treatment of PDP is 34 mg daily. Therefore, the most pertinent comparison of adverse events for the purpose of labeling and review is between placebo and pimavanserin 34 mg daily (PIM 34 mg) in the 6-week controlled trial Parkinson's disease psychosis population (PDP6).

Within the nervous system disorders SOC, the only AE experienced by ≥2% of patients on pimavanserin and more frequently than on placebo was dizziness (PIM 34 mg 4.5% and placebo

4.3%); there was no apparent dose-response relationship across the pimavanserin 8.5 mg, 17 mg, or 34 mg groups for this AE. During the first 6 weeks of open-label treatment among the placebo roll-over patients, the incidence of dizziness was 1.6%.

Within the psychiatric disorders SOC, the most frequent AE experienced by patients in the PDP6 population that was more common in the PIM 34 mg group compared with placebo was confusional state (5.9% and 2.6%, respectively), with a small increase in frequency observed with increasing dose (4.3% with pimavanserin 8.5 mg; 4.9% with 17 mg; and 5.9% with 34 mg).

For AEs ≥2% (for the PIM 34 mg or placebo groups) within the infections and infestations SOC, the incidence of urinary tract infection was 7.4% for PIM 34 mg and 6.9% for placebo.

Within the gastrointestinal disorders SOC, the incidence of nausea was 6.9% for PIM 34 mg and 4.3% for placebo, constipation was 4.5% for PIM 34 mg and 2.6% for placebo, and diarrhea was 2.5% for PIM 34 mg and 1.7% for placebo.

More patients in the All PIM group experienced elevated creatine phosphokinase (2.3%) than in the placebo group (1.3%); however, only 0.5% of patients in the PIM 34 mg group experienced an increase in creatine phosphokinase.

Table 10. Adverse Events Experienced by ≥2% of Pimavanserin-Treated Patients and Greater than Placebo in the PDP Placebo-controlled 6-Week Studies and Partial Data from Open-label Trial -015

MedDRA System Organ Class (SOC) Preferred Term	Placebo (N=231) n (%)	PIM 8.5 mg (N=140) n (%)	PIM 17 mg (N=41) n (%)	PIM 34 mg DB (N=202) n (%)	All PIM (N=383) n (%)	PIM 34 mg OL[a] (N=184) n (%)
Nervous System Disorders						
Dizziness	10 (4.3)	7 (5.0)	1 (2.4)	9 (4.5)	17 (4.4)	3 (1.6)
Psychiatric Disorders						
Confusional state	6 (2.6)	6 (4.3)	2 (4.9)	12 (5.9)	20 (5.2)	3 (1.6)
Gastrointestinal Disorders						
Constipation	10 (4.3)	6 (4.3)	0	14 (6.9)	20 (5.2)	4 (2.2)
Investigations						
Blood creatine phosphokinase increased	3 (1.3)	5 (3.6)	1 (2.4)	3 (1.5)	9 (2.3)	1 (0.5)

Source: Applicant's Table PDP6 2-2.1 and ISS page 164
MedDRA version 15.1 was used to categorize the adverse events.
Only adverse events (AEs) that occurred on or after the first administration of trial drug and before or on last dose date +30 are included.
Patients may have more than one AE per system organ class or preferred term, patients were counted at most once per system organ class and preferred term. Denominators for the percentages were the number of patients in each treatment group.
[a] Includes adverse events only up to Day 72 for patients in ACP-103-015 that were in the placebo treatment group in the core studies ACP-103-012, -014, and -020.
* Met $p<0.05$ level of significance using Fisher's Exact test by comparing the AE rate for each pimavanserin group (except for pimavanserin 34 mg OL) versus Placebo.

Laboratory Parameters Reported as Adverse Events in the PDP Placebo-controlled (6-week) Studies (Population PDP6)

The most commonly reported clinical laboratory value AEs were creatine phosphokinase (CPK) increased with 9 reports (2.3%) in the all PIM group and 3 in placebo (1.3%); no other laboratory values were reported as AEs in more than 1% of pimavanserin-treated patients. None of these laboratory-related events in either treatment group were reported as serious adverse events or led to trial discontinuation for any patient.

There were no patients that met criteria for possible Hy's Law cases (defined as any elevated ALT/AST of >3×ULN, ALP <2×ULN, and associated with an increase in bilirubin ≥2×ULN), a widely used test for potential severe hepatotoxicity.

The mean change from Baseline to Week 6 was similar between pimavanserin- and placebo-treated patients for the following laboratory values: alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, bilirubin, blood urea nitrogen, creatinine, uric acid, creatine kinase, and glucose.

Markedly Abnormal Clinical Laboratory Values

Markedly abnormal electrolyte values were seen sporadically across all arms. The only analyte for which findings were seen consistently across all groups was calcium (and specifically for values <2.1 mmol/L); 7 patients (3.1%) in the placebo group and 13 patients (6.9%) in the pimavanserin 34 mg group. The only other analyte for which >2 patients experienced markedly abnormal findings was potassium (>5.5mmol/L); 4 patients (1.8%) in the placebo group and 11 patients in the All PIM group (3.0%). Almost all other markedly abnormal electrolyte values were seen in just one patient and in all cases were in the placebo or pimavanserin 8.5 mg groups.

Markedly abnormal post-baseline clinical chemistry values (among patients with normal values at baseline) were sporadic and showed no consistent patterns.

A total of 12 patients (4 in placebo [1.7%] and 8 in the All PIM group [2.1%]) had a creatine kinase/phosphokinase value ≥3 times the upper limit of normal (ULN). In general elevations were either present at baseline or returned to normal while the patients continued on trial drug indicating the elevated CKs were not due to a drug effect. No patients in the PDP6, PIM 34 mg group who had normal CK values at baseline developed markedly abnormal CK values compared to two patients in the placebo group.

Vital Signs

Across placebo-controlled studies of pimavanserin, vital sign mean values were similar across all treatment groups. In general, a higher percentage of patients in the placebo group than the pimavanserin 34 mg group had an event of orthostatic hypotension. Review of results for vital signs in the open-label long-term studies revealed no clinically relevant mean changes from baseline.

In the outlier analysis, the proportion of patients who met the criteria for markedly abnormal changes in vital signs was similar for the pimavanserin and placebo groups.

Electrocardiograms (ECGs)

As noted above, the cardiac safety profile of pimavanserin has been evaluated clinically in a thorough QT study in healthy normal volunteers and in both short-term placebo-controlled and long-term open-label studies in PDP patients.

The patient data are consistent with the profile observed in the thorough QT study. Sporadic QTcF values ≥500 ms and change from baseline values ≥60 msec were observed in PDP patients treated with pimavanserin 34 mg; the incidence of these changes was generally similar in the pimavanserin and placebo groups. There were no reports of torsade de pointes or any differences from placebo in the incidence of other adverse reactions associated with delayed ventricular repolarization in studies of pimavanserin, including those in PDP patients.

Immunogenicity

The overall incidence of AEs related to immunogenicity/hypersensitivity in the PDP double blind 6-week studies was 1.6% for the All PIM group (N=383) and 1.3% for the placebo group (N=231). There were no significant risk differences for AEs of immunogenicity for pimavanserin compared to placebo.

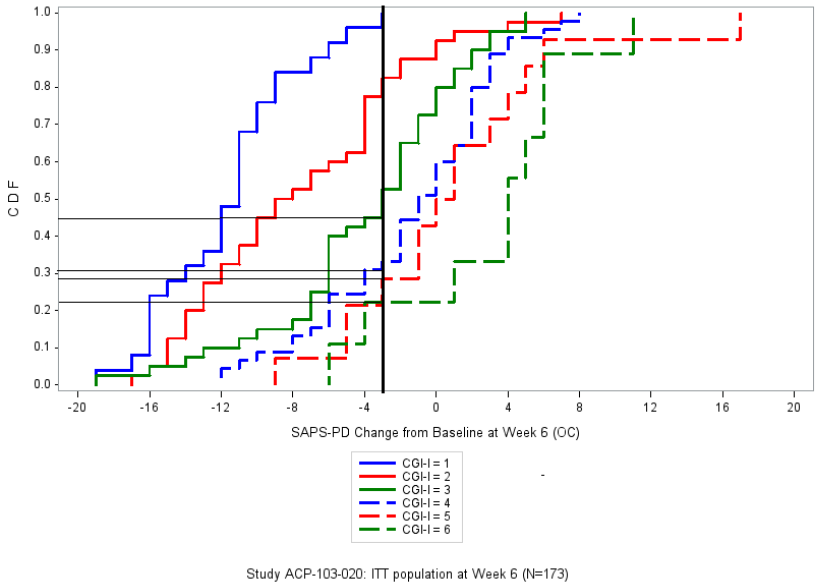
3.4. Discussion of Clinical Meaningfulness

Based on the regression analysis described in a publication on the performance of the SAPS-PD (Voss et al., 2013), a clinically meaningful change, defined as a 1-unit change in the Clinician Global Impression-Improvement (CGI-I) scale, is associated with a 2.33-point change in the SAPS-PD. A 1-unit change on CGI-I is considered a minimally improved intra-patient change on a 7-point CGI-I. The 7 units are: 1 = Very much improved, 2 = Much improved, 3 = Minimally improved, 4 = No Change, 5 = Minimally worse, 6 = Much worse, 7 = Very much worse

A 3-point change in the SAPS-PD for trial ACP-103-020 represents the median of the SAPS-PD change score of the patients who showed minimal improvement (i.e., CGI-I=3) from baseline to Week 6 based on CGI-I assessed at Week 6. The median SAPS-PD change score of the patients rated as much improvement from baseline to Week 6 (CGI-I=2) is 7 points, as shown in the cumulative distribution function (CDF) curves below (Figure 1). The CDF curves also show that there is little separation between minimal improvement (CGI-I =3), no change (CGI-I=4), minimally worse (CGI-I=5), and much worse (CGI-I=6). They show that large percentages of "no change" and "worsened" patients also had ≥ 3-point change in SAPS-PD (i.e., 44%, 31%, 29%, and 22% for minimally improved, no change, minimally worse, and much worse, respectively). That is, there is a certain amount of "noise" or uncertainty associated with using the 3-point change as the threshold for clinical meaningfulness. In this regard, a larger threshold

that represents clinically meaningful improvement with higher certainty, such as 7-point or 5-point change, may be considered.

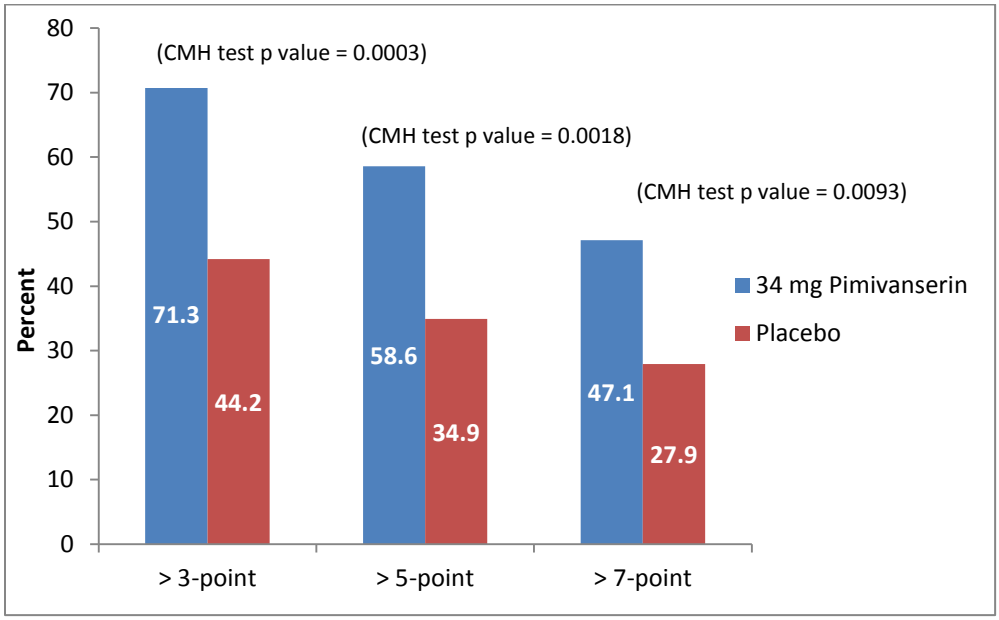
Figure 1. SAPS-PD by CGI level



Source: CDER Office of Biostatistics analysis

The histogram below (Figure 2) shows the percentages of patients meeting the 3-, 5-, and 7-point changes in the two treatment arms. For all levels of improvement, the pimavanserin group was more likely to show a change. However, the histogram shows that the 3-point threshold may be a low estimate of clinically meaningful change given that 44.2 % of the patients in the placebo arm had more than a 3-point change in the SAPS-PD total score. On the other hand, a 5 to 7-point change may be a reasonable threshold for clinically meaningful improvement given that fewer patients in the placebo arm attain that level of improvement.

Figure 2. Proportion of Patients with SAPS-PD Score Improvement at the End of Week 6



Source: CDER Office of Biostatistics analysis

A 50% reduction in the SAPS-PD corresponds to what was demonstrated by Leucht, et al., as "much improved" (but still less than "very much improved"). A 30% improvement corresponds to a "minimal improvement."

The following table lists the number of patients who experienced a 50% reduction in the SAPS-PD score during the clinical trials.

Table 11. SAPS-PD Responder Rate Analysis by Treatment and Visit in Trial 103-020

Treatment Group	Responder rate (Number of Patients who had a response, divided by total Number of Patients)	
	Week 2	Week 6
Pimavanserin	20.2% (19/94)	**37.2% (35/94)**
Placebo	20.0% (18/90)	**27.8% (25/90)**
Both groups	20.1% (37/184)	32.6% (60/184)

Note: A responder is defined as having >50% improvement from baseline and with the assumption that all dropouts are non-responders.

Calculating a number needed to treat (NNT)

$$NNT = \frac{1}{(IMPact/TOTact) - (IMPcon/TOTcon)} = \frac{1}{35/94 - 25/90} = 11$$

Therefore, one must treat 11 patients for one patient to receive a 50% reduction in the SAPS-PD. Calculating a number needed to harm (NNH) provides the balance to the absolute chance of efficacy. Using any serious adverse event, including death, as the definition of "harm," 16 of 202 Pimavanserin treated patients experienced a SAE in the PDP6 population; 8/231 placebo treated patients experienced a SAE in the PDP6 population.

$$NNH = \frac{1}{(IMPact/TOTact) - (IMPcon/TOTcon)} = \frac{1}{16/202 - 8/231} = 22$$

The NNH/NNT ratio is 2. Put another way, for every 2 patients who achieve a 50% reduction in the SAPS-PD, 1 patient will experience a serious adverse event that is attributable to pimavanserin.

3.5. Nonclinical Drug Toxicity Profile

Pimavanserin is a cationic amphiphilic drug (CAD). CADs are known to cause phospholipidosis (PLD), the excessive accumulation of phospholipids in cells, in animals and humans. Many marketed drugs are CADs and cause drug-induced PLD in animals and humans (e.g., fluoxetine, choroquine, amiodarone). PLD is usually reversible after cessation of drug treatment; however, high or prolonged exposures to CADs may lead to dose-limiting functional and structural tissue damage (e.g., nephrotoxicity, pulmonary toxicity, myopathy and retinopathy) (Halliwell, 1997; Vonderfecht, et. al., 2004).

In the case of pimavanserin, multi-organ "systemic" PLD was observed in mice, rats and monkeys which were both dose- and duration-dependent and observed as early as after 14 days of daily administration. The number of tissues/organs affected in mice, monkeys and rats was extensive at 5, 15 and over 30, respectively, with the lungs and kidneys being the most severely affected. Because multi-organ PLD was observed across all animal species tested, it is reasonable to assume that the PLD-related toxicity of chronic inflammation with secondary fibrosis in the lungs can also occur across species and is dependent on the severity of the PLD and duration of drug exposure. Multi-organ PLD with chronic inflammation and fibrosis and PLD-related morbidity/mortality is a clinically relevant finding.

There is 9-fold safety margin, based on AUC, to the No Observed Effect Level (NOEL) for chronic inflammation with secondary fibrosis in the lungs, and PLD-related morbidity/mortality in rats compared to the maximum recommended human dose (MRHD) of 34 mg/day pimavanserin. There is a 5-fold safety margin to the NOEL for PLD leading to chronic inflammation in the lungs of rats. This margin is acceptable from a nonclinical perspective for the indication of Parkinson's disease psychosis. This is because the average life expectancy for these patients is not more than a few years; therefore, there is less concern regarding the possibility of developing multi-organ PLD that may lead to chronic inflammation and possible secondary fibrosis in the lungs. Concern would be greater in a patient population where life expectancy is longer.

Potential clinical manifestations of phospholipidosis: No events suggestive of hepatocellular changes were reported in the PDP6 Population. The most frequent respiratory event was dyspnea (0.8% All PIM, no placebo patients, and 1.6% pimavanserin 34 mg open-label). For renal events, only 1 patient in the pimavanserin 34 mg open-label group experienced an AE of acute renal failure, whereas, across all other treatment groups, no kidney-related events were reported.

4 SUMMARY

Pimavanserin is a selective serotonin inverse agonist under review for the treatment of psychosis associated with Parkinson's disease (PDP). Unlike other antipsychotics, pimavanserin does not produce dopamine blockade; therefore, it has the potential to reduce symptoms of hallucinations, delusions, and agitation without adversely affecting the motor symptoms of Parkinson's disease. PDP is a disabling condition that ultimately affects half of the millions who suffer from Parkinson's disease worldwide. There are no FDA-approved drug treatments for this condition.

The primary clinical outcome variable to establish efficacy of pimavanserin was the 9-item, Schedule for the Assessment of Positive Symptoms - Parkinson's Disease (SAPS-PD) scale. This is the first use of this scale in a clinical trial. The mean difference in change from baseline on the SAPS-PD for pimavanserin-treated patients in the 6-week trial was 3.1 points (p=0.001) better than the change for placebo-treated patients; this represents an improvement in the measured psychotic symptoms of 23.1% over placebo. Linking the change in rating scales of psychosis to the CGI, Leucht finds that a 22-34% improvement in scales that measure psychotic symptoms correlates to a CGI score of "minimally improved" (Leucht, et al., 2006). Thus, the highly statistically significant treatment difference from placebo demonstrated in the single positive clinical trial (ACP-103-020) appears quite modest, on average, although the benefit received by some patients was more than minimal (see Figure 2). In any case, the benefit observed in this trial must be weighed against the potential harms of the drug.

The observed risk for serious adverse events including death in the 6-week, placebo-controlled trial (PDP6) population for the development of pimavanserin is 2.38 (95% CI 1.00 to 5.73, p=0.05) for 34 mg vs. placebo. SAEs occurred in 16/202 (7.9%) patients taking pimavanserin 34 mg versus 8/231 (3.5%) placebo treated patients in the PDP6 population. No individual SAE appeared to dominate this difference and there was no unifying pathological mechanism or premonitory signal. In the long-term PDP open-label treatment population, there were 51 deaths among 459 treated patients with PDP (11.1%).

5 REFERENCE LIST

Andreasen NC. Scale for the Assessment of Positive Symptoms. Iowa City: University of Iowa Press; 1984.

Goetz, C. G., and G. T. Stebbins, 1995, Mortality and hallucinations in nursing home patients with advanced Parkinson's disease: Neurology, v. 45, p. 669-71.

Halliwell, H.W. (1997) Cationic Amphiphilic Drug-Induced Phospholipidosis. *Tox. Path.*, 25: 53-60.

Ikebe, S., T. Harada, T. Hashimoto, I. Kanazawa, S. Kuno, Y. Mizuno, E. Mizuta, M. Murata, T. Nagatsu, S. Nakamura, H. Takubo, N. Yanagisawa, and H. Narabayashi, 2003, Prevention and treatment of malignant syndrome in Parkinson's disease: a consensus statement of the malignant syndrome research group: Parkinsonism Relat Disord, v. 9 Suppl 1, p. S47-9.

Leucht, S., J. M. Kane, E. Etschel, W. Kissling, J. Hamann, and R. R. Engel, 2006, Linking the PANSS, BPRS, and CGI: clinical implications: Neuropsychopharmacology, v. 31, p. 2318-25.

Levy G, Tang M-X, Louis ED, et al. The association of incident dementia with mortality in PD. Neurology 2002;59:1708–13

Mizuno, Y., H. Takubo, E. Mizuta, and S. Kuno, 2003, Malignant syndrome in Parkinson's disease: concept and review of the literature: Parkinsonism Relat Disord, v. 9 Suppl 1, p. S3-9.

Ogawa, E., R. Sakakibara, M. Kishi, and F. Tateno, 2012, Constipation triggered the malignant syndrome in Parkinson's disease: Neurol Sci, v. 33, p. 347-50.

Vonderfecht, S.L, et al., (2004); Myopathy related to administration of cationic amphiphilic drug and the use of multidose drug distribution analysis to predict its occurrence. Tox. Path. 32: 318-325.

Voss T, Bahr D, Cummings J, et al. Performance of a shortened Scale for Assessment of Positive Symptoms for Parkinson’s disease psychosis. Parkinsonism and Related Disorders. 2013; 19:295-299.

Williams, D. R., and A. J. Lees, 2005, Visual hallucinations in the diagnosis of idiopathic Parkinson's disease: a retrospective autopsy study: Lancet Neurol, v. 4, p. 605-10.

6 APPENDICES

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDA 207,318
Drug Name:	Pimavanserin (Nuplazid™)
Indication(s):	Psychosis associated with Parkinson's Disease
Applicant:	Acadia Pharmaceuticals Inc.
Date(s):	September 1, 2015 (Submission Stamp date);
	May 1, 2016 (PDUFA due date)
Review Priority:	Priority
Biometrics Division:	Division of Biometrics I
Statistical Reviewer:	Eiji Ishida, M.S.
Concurring Reviewers:	Peiling Yang, Ph.D.; H. M. James Hung, Ph.D.
Medical Division:	DPP (Division of Psychiatry Products)
Clinical Team:	Paul Andreason, M.D.
Project Manager:	Brendan Muoio, Pharm.D.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

The single pivotal study (Study ACP-103-020) has established statistical evidence that pimavanserin 40 mg is efficacious as a treatment of Parkinson's Disease Psychosis (PDP) and does not worsen Parkinson's Disease status. No major issue that may affect the main statistical conclusions was found.

The primary efficacy endpoint was the change from baseline to Week 6 in score of a novel instrument SAPS-PD (Scale for the Assessment of Positive Symptoms for Parkinson's Disease). The SAPS-PD assesses Parkinson's Disease Psychosis (PDP) for Hallucination and Delusion symptoms. Based on SAPS-PD, pimavanserin 40 mg has been shown to be more efficacious than placebo for an acute treatment of PDP.

The key secondary endpoint was the change from baseline to Week 6 in UPDRS Parts II and III Combined score, the respective part of which assesses activities in daily life and motor functions of Parkinson's Disease patients. A non-inferiority of pimavanserin to placebo has been concluded based on a non-inferiority test with a non-inferiority margin of 5 points. To be specific, pimavanserin ruled out >2.72 points worse than placebo based on the 95% confidence interval of the treatment effect relative to placebo.

A Psychopharmacologic Drugs Advisory Committee (PDAC) meeting will be held on March 29, 2016. One of the questions addressed to the PDAC may likely be about whether the treatment effect observed in this study, 3 points in SAPS-PD as an observed effect (a difference of pimavanserin from placebo in least square mean estimate), is clinically meaningful. This review may provide information on clinical effectiveness of the candidate new treatment.

2 INTRODUCTION

Acadia Pharmaceuticals Inc. filed a new drug application (NDA) of pimavanserin for an indication of Psychosis associated with Parkinson's Disease (PDP). The candidate drug is pimavanserin 40 mg. In the NDA filing, the sponsor included one pivotal study and three supportive studies for evaluation of efficacy and safety. At Type C Meeting, held on April 9, 2013, the FDA agreed to the sponsor's plan to file the NDA based on data from the single, positive study (ACP-103-020) and supportive data from other studies (see Table 1).

The pivotal study Study ACP-103-020, titled "A Multi-Center, Placebo-Controlled, Double-Blind Trial to Examine the Safety and Efficacy of Pimavanserin in the Treatment of Psychosis in Parkinson's Disease", was conducted in US and Canada with 199 randomized patients. The sponsor evaluated efficacy of a fixed dose of 40 mg pimavanserin compared to placebo in the 6-week dosing duration. The sponsor claims this study established pimavanserin efficacy for the PDP indication.

A supportive study, Study ACP-103-012, conducted in US, Europe (mostly Eastern Europe) and India with 298 randomized patients, had two fixed doses of pimavanserin 10 mg and 40 mg to be compared to placebo in the 6-week dosing duration. The sponsor was unable to establish efficacy of pimavanserin in Study ACP-103-012 (Refer to Appendix A3 for this reviewer's brief summary of efficacy results). Another supportive study, Study ACP-103-014, had two fixed doses of pimavanserin 10 mg and 20 mg compared to placebo in the 6-week dosing duration. This study was early terminated when 123 patients of the planned 280 patients were enrolled. The third supportive study, Study ACP-103-006, conducted in US with 60 randomized patients, was a flexible dose placebo-controlled study (20 mg, 40 mg and 60 mg) in the 4-week dosing duration.

The present statistical review evaluates efficacy of pimavanserin for the labeling claim, based on the single, pivotal study (Study ACP-103-020). It is noted that at the planning stage of the pivotal study, the agency required that the sponsor evaluate motor symptoms of Parkinson's Disease in their planned antipsychotic efficacy studies of pimavanserin.

2.1 Overview

Study ACP-103-020 was a multicenter, randomized, fixed-dose, double-blind, placebo-controlled efficacy study in which 199 subjects were randomized across 66 centers (63 in the US and 3 in Canada). The sponsor evaluated efficacy of a fixed dose of 40 mg pimavanserin compared to placebo in the 6-week dosing duration. The sponsor defined the ITT population (later renamed as 'mITT') as consisting of randomized patients who had their baseline score and at least one post-baseline score of the primary efficacy measure. The number of subjects of the mITT population was 185 (95 patients for pimavanserin and 90 patients for placebo).

According to the sponsor, the study population has been largely the same for all pimavanserin studies for the PDP indication; the entry criteria have been consistent with the NINDS/NIMH

established diagnostic criteria for PDP, as published by a Movement Disorder Society task force in 2007. However, according to the sponsor, the entry criteria for severity/frequency of psychosis of Study ACP103-020 are higher than for those of Studies ACP-103-012 and -014 (See Table 1). Specifically, in Study ACP-103-020, the study patients had to meet the condition of "A total NPI-H+D score ≥6 OR a score of ≥4 on either the H or D domain", while in Study ACP-103-012 that of "A score of ≥4 on either the H or D domain."

Table 1: Major Study Elements of Pimavanserin Placebo-Controlled Phase III Studies for PDP

Study	**ACP-103-020**	**ACP-103-012**	**ACP-103-014**
	Pivotal	Supportive	
N = # randomized patients	N = 199	N = 298	N = 123 (The planned N was 280, but was terminated early)
Dosing Duration (weeks)	6	6	6
Dose (daily)	Fixed dose 40 mg	Fixed doses 10 mg and 40 mg	Fixed doses 10 mg and 20 mg
Primary Efficacy Endpoint [Hypothesis test result]	Change from Baseline in SAPS-PD [Significant]	Change from Baseline in SAPS H+D [Not Significant]	Change from Baseline in SAPS H+D [N/A]
Key Secondary Endpoint [Hypothesis test result]	Change from Baseline in UPDRS II+III [Non-inferiority established]	Change from Baseline in UPDRS II+III [Not tested due to non-significant result in primary efficacy]	Change from Baseline in UPDRS II+III [N/A]
Region	US and Canada	US, Europe (Mostly Eastern) and India	US and Europe
Major difference in Study Population	**NPI-H+D: Entry Criteria at Screening**		
	A total NPI-H+D score ≥6 OR a score of ≥4 on either the H or D domain	A score of ≥4 on either the H or D domain	
	SAPS-PD: Entry Criteria at Baseline		
	A SAPS global item score (global H or D) of ≥3 and a score of ≥3 on at least one other non-global item	N/A	

NPI=Neuropsychiatric Inventory; SAPS= Scale for the Assessment of Positive Symptoms; H=Hallucinations; D=Delusions; PD=Parkinson's Disease
Source: Reviewer's documents based on CSR, Study Protocol and SAP.

The primary efficacy measure of Study ACP-103-020 was SAPS-PD[1], a scale the sponsor proposed at Type C Meeting with the FDA held on April 26, 2010. This scale was developed based on SAPS (Scale for the Assessment of Positive Symptoms) as a novel instrument to assess Parkinson's Disease Psychosis (PDP). As shown in Table 1, the primary efficacy measure used

[1] Fernandez, et. al (2008) Movement Disorders, Vol.23 No4. 2008. Pp484-500

in Studies ACP-103-012 and -014 was SAPS-H+D, which is a 20-item scale for assessing Hallucinations and Delusions from SAPS. SAPS-PD consists of 9 items selected from the 20 items of SAPS-H+D.

The FDA required the sponsor to adequately assess for worsening of Parkinson's Disease in any PDP study of ACP-103 (pimavanserin), and that the sponsor formally conduct a non-inferiority test to demonstrate that pimavanserin is not inferior to placebo on some measure of PD status. The sponsor chose UPDRS II+III (UPDRS Part II (activities of daily living) and Part III (motor examination)) scores to pre-specify the required non-inferiority test. (See Section 3.2.2 for the selected non-inferiority margin).

The sponsor's Clinical Study Report (CSR) of Study ACP-103-020 indicates that an estimated pimavanserin treatment effect (relative to placebo in mean change from baseline score of SAPS-PD) was 3.06, and that a statistically significant difference (pimavanserine's superiority to placebo) was established at the p-value of 0.0014. It also indicates that non-inferiority of pimavanserin to placebo on Parkinson's Disease status was concluded based on the combined score of the UPDRS II+III.

2.2 Data Sources

The pimavanserin NDA submission is located at the FDA server:

\\Cdsesub1\evsprod\NDA207318\0000

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

The sponsor submitted all NDA efficacy study data in the original submission. The sponsor collected clinical study data of Study ACP103-020 in their legacy database (raw data), and generated legacy analysis data from the raw data. Prior to the NDA submission, the sponsor converted the legacy data to SDTM datasets. The submission of Study ACP103-020 contained the raw datasets (legacy data), and their converted legacy analysis datasets and CDISC SDTM datasets. This reviewer noticed that treatment assignment information was not included in the raw data but only in the legacy analysis datasets and SDTM datasets, and found through the communications with the sponsor that the randomization schedule and the raw data were directly merged into legacy analysis datasets and SDTM data sets. This reviewer confirmed that the planned treatment assignments of the randomization schedule match the actual treatment assignments of the legacy analysis datasets and SDTM datasets.

Given the treatment assignments incorporated into the submitted raw data, this reviewer verified that the analysis results the sponsor obtained from the legacy analysis data are consistent with the raw data that the treatment assignments of the randomization schedule are incorporated into.

3.2 Evaluation of Efficacy

The primary objective of this study was to demonstrate the antipsychotic efficacy of pimavanserin in subjects with Parkinson's Disease Psychosis (PDP) as measured by a decrease in the severity and/or frequency of hallucinations and/or delusions. One key secondary endpoint was pre-specified to demonstrate that pimavanserin does not worsen motor symptoms of Parkinson's Disease (PD) in PDP subjects; for this purpose a non-inferiority test was planned based on the UPDRS II+III combined score. The multiple test procedure for the primary efficacy and key secondary endpoints was agreed upon with the FDA.

3.2.1 Study Population and Baseline Severity/Frequency of Psychosis

General Characteristics of Study Population

The study subjects' psychotic symptoms developed after the diagnosis of Parkinson's disease was established. These symptoms must include visual hallucinations and/or auditory hallucinations, and/or delusions. Their psychotic symptoms had to be present for at least one month. Study subjects actively experienced psychotic symptoms each week during the month prior to the Screening visit (Inclusion criteria 3 and 4).

Severity of Study Population (Table 2 and Table 3)

The study subjects' symptoms had to be severe enough to warrant treatment with an antipsychotic agent, which were documented at screening by items A (Delusions) and B (Hallucinations) of the NPI[2], referred as NPI-H+D. A score of 4 or greater on either the Hallucinations (Frequency x Severity) or Delusions (Frequency x Severity) scales OR a total combined score of 6 or greater was required (Inclusion criteria 5).

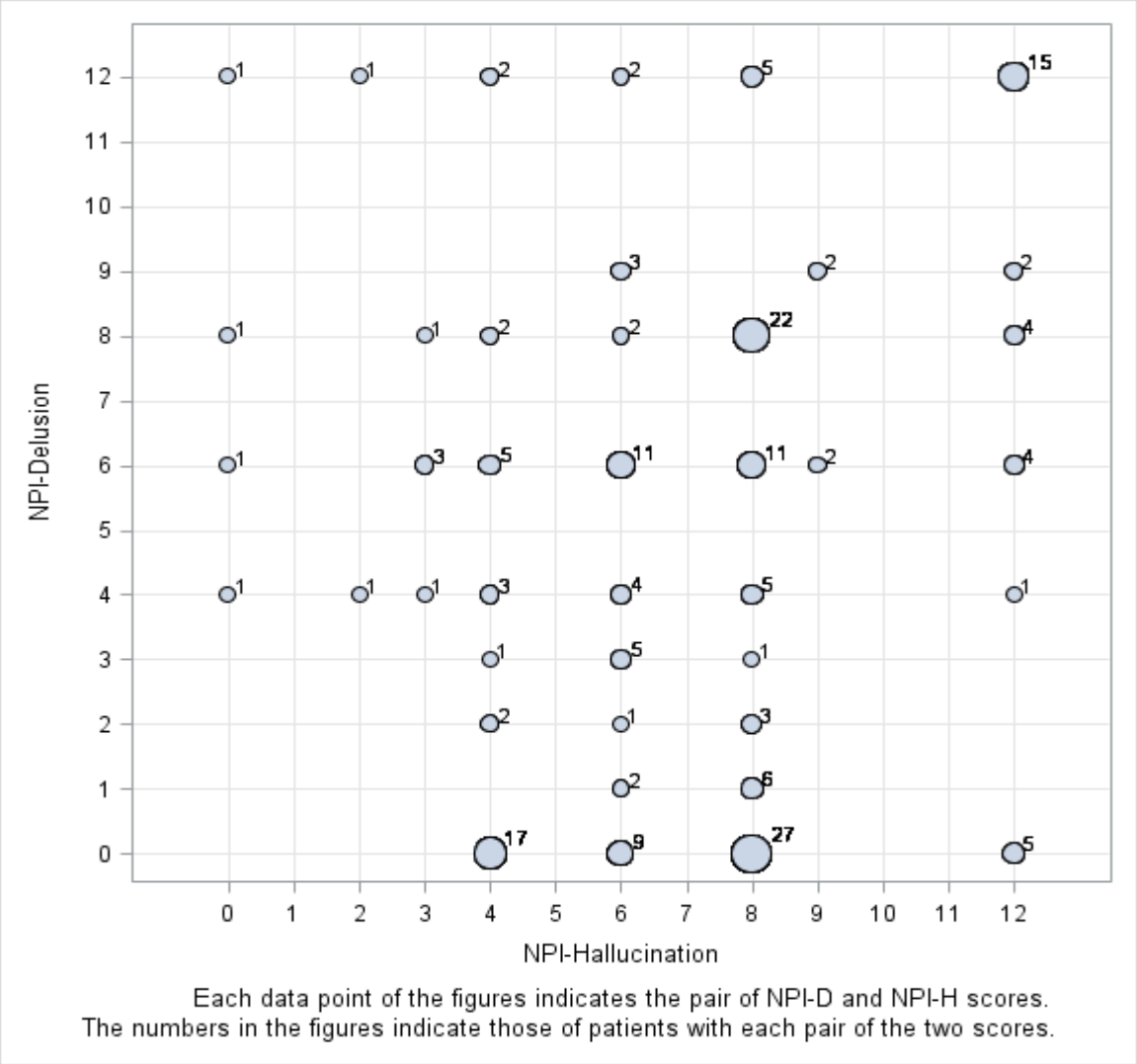
In addition, at the baseline visit, study subjects must have met a criterion based on SAPS-PD: a SAPS Hallucinations or Delusions global item (H7 or D13) score $\geq$3 AND a score $\geq$3 on at least one other non-global item (Inclusion criteria 6). The items A and B of the NPI and SAPS-PD are provided in this review (see Appendix).

The two tables (Table 2 and Table 3) display the severity of all randomized patients based on NPI-H+D (or NPI A and B) and the baseline severity of mITT patients by treatment group based on the SAPS-PD total score. Both mean NPI-H+D total score and mean SAPS-PD score are almost the same for the two treatment groups.

Out of 185 mITT patients at baseline, 58 patients had a NPI-Delusion score of zero (Figure 1) and 33 patients had a SAPS-PD Delusion score of zero (Figure 2), while 4 patients had a score of zero in both NPI-Hallucination and SAPS-PD Hallucination as seen in both figures. Thus, the study patients may have been less severe in their delusion symptoms. This fact may also be seen in the lower mean score of NPI-Delusion of 4.84 in contrast with that of 7.21 in NPI-Hallucination. A similar observation may be made for the SAPS-PD Delusion and Hallucination scores (Figure 2).

[2] NPI: Neuropsychiatric Inventory

Figure 1: Distribution of NPI-H+D by Hallucination/Delusion at Screening (185 mITT Patients)



Note: The numbers of the plots indicate those of subjects.
Source: Reviewer's analysis

Table 2: NPI-H+D Score at Screening and at Baseline by Treatment Group

Screening entry criteria for NPI: A total NPI-H+D score ≥6 OR A score of ≥4 on either the H or D domain	NPI-H+D Screening/Baseline Score		
	Screened patients (N=199)	mITT Population (N=185)	
	Prior to randomization	Pimavanserin (N=95)	Placebo (N=90)
Mean NPI-H+D Delusion score (SD)	4.84 (4.14)	4.8 (4.21)	4.9 (4.09)
Mean NPI-H+D Hallucination score (SD)	7.21 (2.80)	7.1 (2.81)	7.3 (2.83)
Mean NPI-H+D Total score (SD)	12.06 (5.58)	11.8 (5.85)	12.2 (5.33)

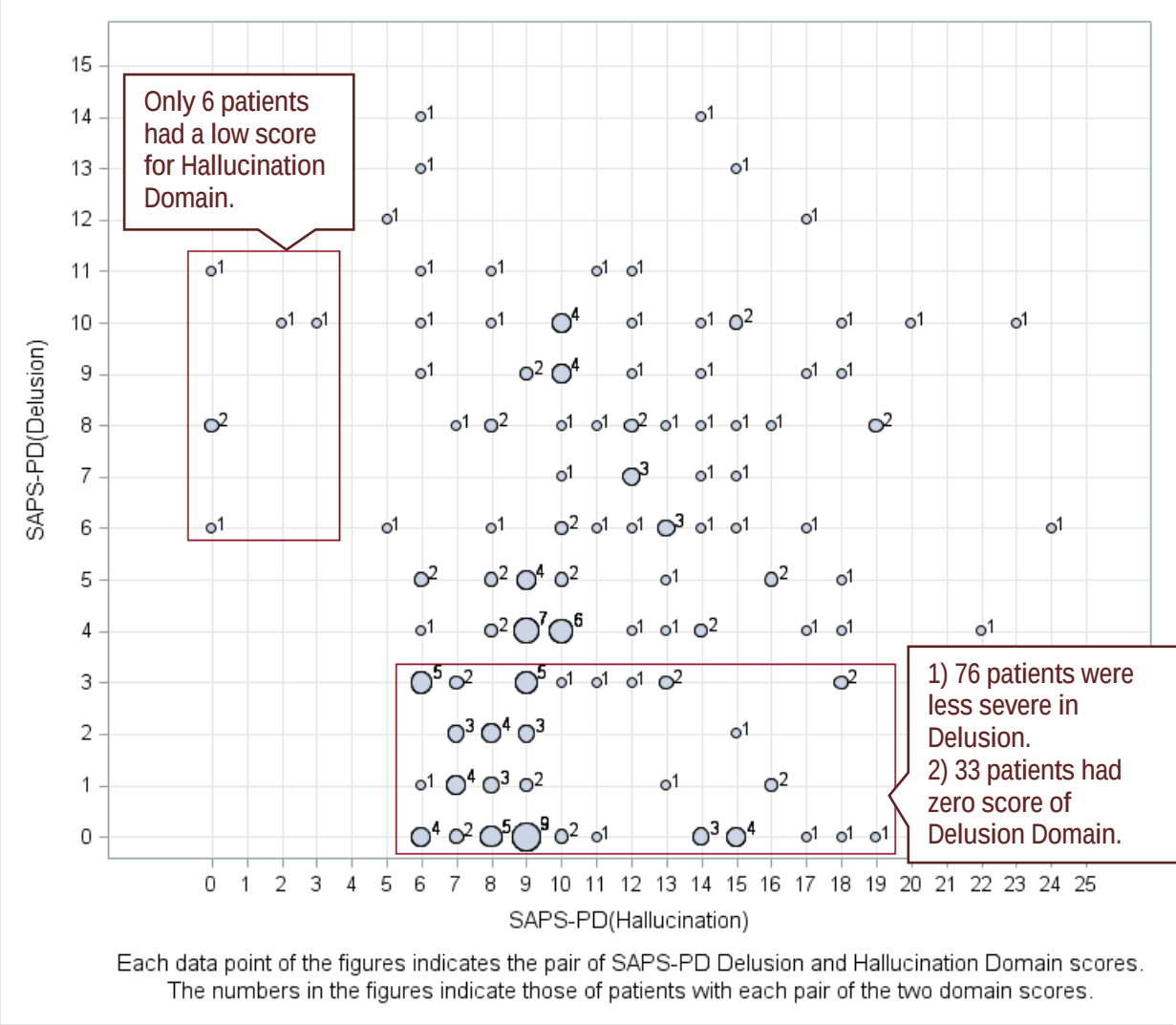
SD denotes standard deviation.
Source: Reviewer's analysis

Table 3: SAPS-PD Baseline Score of mITT Patients (Total Score and Domain Scores)

Randomization entry criteria for SAPS-PD: A SAPS global item score (global H or D) of ≥3 AND A score of ≥3 on at least one other non-global item	SAPS-PD Baseline Score - mITT Population (N=185)	
	Pimavanserin (N=95)	Placebo (N=90)
Mean SAPS-PD Delusion score (SD)	4.76 (3.59)	4.78 (3.82)
Mean SAPS-PD Hallucination score (SD)	11.13 (4.58)	9.96 (3.80)
Mean SAPS-PD Total score (SD)	15.88 (6.12)	14.73 (5.55)

SD denotes standard deviation.
Source: Reviewer's analysis

Figure 2: Distribution of SAPS-PD Hallucination and Delusion Domain Scores at Baseline (N=185)



Note: The numbers of the plots indicate those of subjects.
Source: Reviewer's analysis

3.2.2 Study Design and Endpoints

Study Design

The pivotal study, Study ACP-103-020, was a six-week, multi-center, randomized, double-blind, placebo-controlled study. The study was conducted on an outpatient basis. The baseline disease severity was characterized based on NPI-H+D[3], assessed at the screening visit and SAPS-PD (the modified 9-item SAPS[4] for Hallucinations and Delusions) assessed at the baseline (pre-randomization) visit. A total of the planned sample size was 200, and the randomization was performed at an equal ratio for pimavanserin or placebo group. The planned efficacy assessments were scheduled to be performed at Day 15 (Week 2), Day 29 (Week 4) and Day 43 (Week 6). There was only one planned post-baseline assessment for UPDRS Parts II and III to evaluate motor functions, which was on Day 43 (Week 6). A follow-up visit (Day 71) was scheduled 4 weeks after the last regular study visit for those subjects who do not continue into an open-label extension protocol.

Primary Efficacy Endpoint

The primary efficacy endpoint was pre-specified to be the mean change from pre-dose baseline (Day 1) at Week 6 in SAPS-PD, consisting of Hallucination and Delusion scales. This endpoint assesses an improvement in hallucination and delusion symptoms of Parkinson's Disease as one clinical entity over the 6-week treatment duration.

The sponsor planned to have a Type C Meeting (a teleconference scheduled to be held on April 26, 2010) and had the following question (Question 1) in the meeting package:

Does the Agency agree that modification of the primary endpoint based on a 9-item subset of the SAPS-H+D scale is supported by the baseline (pre-treatment) data from pimavanserin and clozapine studies and is appropriate for the planned Phase 3 trial?

The FDA's preliminary comment was as follows:

Yes, the modified primary efficacy endpoint is acceptable. We agree that the 9-item subset of the SAPS-H+D scale appears to be supported by the factor analysis on the baseline data from the studies. The modified primary endpoint appears to have improved clinical relevance and face validity, compared to the 20-item scale.

The sponsor cancelled the teleconference, determining that receiving the Agency's preliminary comments were sufficient for them.[5]

[3] NPI: Neuropsychiatric Inventory- Hallucination and Delusion
[4] SAPS: Scale for the Assessment of Positive Symptoms
[5] DARRTS: Meeting Cancellation Form, dated April 28, 2010. (IND68384)

Key Secondary Endpoint

The key secondary endpoint is the mean change from Baseline in the combined UPDRS Part II and III scores (UPDRS-II+III) at Week 6.

The non-inferiority margin in the UPDRS Parts II and III Combined Score was discussed with the FDA during the 29 June 2006 Type C teleconference[6]. It was agreed that a change of approximately 5 points or greater is currently considered to be the minimal clinically important change in the UPDRS motor score.

Sample Size Estimation

The following is the sponsor's description of the sample size calculation:
"The primary endpoint is the difference in the mean absolute change in the combined scores for the modified 9-item SAPS (Scale for the Assessment of Positive Symptoms) Hallucinations and Delusions domains (SAPS-PD) between the active arm and the control arm from Day 1 (Baseline) to Day 43 (Week 6).

Assuming a clinically meaningful treatment difference in the mean change in the modified 9-item SAPS-PD Scores of 3.00 from Day 1 (Baseline) to Day 43 (Week 6) and using an estimated standard deviation of 6.5, a total of 100 subjects per treatment arm is required to achieve a 90% chance that the comparison (i.e., 40 mg pimavanserin versus placebo) will be significant at the alpha level of 0.05 for a 2-sided t-test.

As currently planned, the Phase III study will also be powered to detect a 5-point margin of difference from placebo on the key secondary endpoint, UPDRS (Unified Parkinson's Disease Rating Scale) Parts II and III. In a similarly designed study ACP-103-012, at the end of the 6-week study period, the standard deviations of UPDRS change from Baseline were 9.25 and 9.42 points in the placebo and pimavanserin 40 mg groups, respectively. Assuming a common standard deviation of 9.4 points, and conservatively assuming that the mean change from Baseline in UPDRS at Day 43 (Week 6) in the pimavanserin 40 mg group is 0.5-point greater (worse) than the placebo group, 100 subjects per treatment group will have a 92% power to demonstrate that the pimavanserin 40 mg group is not inferior to the placebo group within a margin of 5 points at a 1-sided 0.025 significance level."

3.2.3 Statistical Methodologies

3.2.3.1 Primary Efficacy

Primary Analysis for Primary Efficacy

Mean change from baseline in SAPS-PD was analyzed using a restricted maximum likelihood (REML)-based mixed model repeated measures (MMRM) method. The model included the fixed, categorical effects of treatment (placebo or pimavanserin; 2 levels), visit (Weeks 2, 4 and 6; 3 levels) and treatment-by-visit interaction, as well as the continuous, fixed covariate of

[6] DARRTS: The meeting minutes is dated 10 July 2006. (IND68384)

baseline score. An unstructured covariance matrix was used to model the within-subject errors. The Kenward-Roger approximation was used to estimate denominator degrees of freedom. The significance test was based on least-square means comparison at the Week 6 using a 2-sided $\alpha = 0.05$. The primary analysis used observed measurements alone and no missing data were imputed.

Sensitivity Analysis for Primary Efficacy

In the primary efficacy analysis, missing data are assumed to follow the same distribution as observed data. Since this assumption may not be true, it may be important to evaluate efficacy under a scenario that missing data have values that are less efficacious than the observed data. The sponsor pre-specified a few sensitivity analyses, two of which may be important to mention in this review: ANCOVA with the WOCF (worst observation carried forward) imputation, which assumes that the missing data may be set equal to the worst observation of the patient and ANCOVA with the LOCF (last observation carried forward) imputation, which assumes that the missing data may be set equal to the last observation of the patient.

3.2.3.2 Key Secondary Endpoint and Analysis

The key secondary endpoint was the mean change from baseline to Week 6 in the combined score of the UPDRS Part II (activities of daily living) and Part III (motor examination). There was only one post-baseline assessment, and the primary analysis was the ANCOVA analysis (OC, i.e., completers only) with the mITT analysis set. To control the study-wise type I error rate, the test for key secondary endpoint was planned to be performed only when the null hypothesis for the primary efficacy endpoint was rejected at the significance level of 0.05 (two-sided).

3.2.3.3 Some Other Secondary/Supportive Endpoints and Analysis

Clinical Global Impressions - Severity (CGI-S) and - Improvement (CGI-I)

The sponsor planned to analyze the difference between pimavanserin versus placebo in Least Square mean changes from baseline (CGI-S) and in Least Square mean score at Week 6 (CGI-I) using MMRM as well as ANCOVA with LOCF and OC in the mITT analysis set.

SAPS-H+D (20-Item) Score

The sponsor planned to analyze the difference between pimavanserin versus placebo in SAPS-H+D Least Square mean changes from baseline using MMRM as well as ANCOVA with LOCF and OC in the mITT analysis set.

Combined SAPS Hallucinations and Delusions Global Rating of Severity (GSAPS-H+D)

The sponsor planned to analyze the difference between pimavanserin versus placebo in GSAPS-H+D LS-mean changes from baseline will be analyzed using MMRM as well as ANCOVA with LOCF and OC in the mITT analysis set.

SAPS Subscales and Global Rating Score

The sponsor planned to analyze the SAPS Hallucinations domain score (SAPS-H) and Delusions domain score (SAPSD), which is based on SAPS-H+D and the global rating severity scores

within each sub-scale (GSAPS-H, GSAPS-D). The analysis was conducted using MMRM as well as ANCOVA with LOCF and OC in the mITT analysis set.

3.2.3.4 Center Effects

More than 50 centers participated in this study and many centers had a small number of subjects. However, as the primary efficacy endpoint was rated using a centralized rater service, study center was not pooled. The primary analysis did not include center effect as a factor in its analysis model.

3.2.4 Patient Disposition, Demographic and Baseline Characteristics

Demographic Characteristics (Table 4)

The summary statistics of the study demographic characteristics are provided in the table below.

Table 4: Demographic Characteristics

mITT population	Placebo (N=90)	Pimavanserin (N=95)	Total (N=185)
Gender n (%)			
Male	52 (57.8)	64 (67.4)	116 (62.7)
Female	38 (42.2)	31 (32.6)	69 (37.3)
Age (years)			
Mean (SD)	72.4 (7.92)	72.4 (6.55)	72.4 (7.23)
Age Categories n (%)			
<65 Years	11 (12.2)	11 (11.6)	22 (11.9)
65-75 years	50 (55.6)	53 (55.8)	103 (55.7)
>75 Years	29 (32.2)	31 (32.6)	60 (32.4)
Race Category n (%)			
White	85 (94.4)	90 (94.7)	175 (94.6)
Non-White	5 (5.6)	5 (5.3)	10 (5.4)

SD denotes standard deviation.
Source: Table 14.1.2.1.1 of the CSR

Randomization and Primary Analysis Set

Randomization: A total of 199 eligible patients were randomized to pimavanserin or placebo at an equal ratio (N=94 for placebo and N=105 for pimavanserin).

Primary Analysis Set: The sponsor's primary analysis set was the modified intent-to-treat (mITT) population defined as a set of randomized patients who have their baseline score of SAPS-PD and at least one post-baseline score of SAPS-PD. The sponsor excluded 14 of 199 randomized patients from their mITT analysis population[7]. This reviewer verified that the removal of the excluded patients from the mITT analysis for primary efficacy does not favor pimavanserin for its efficacy evaluation. The sponsor's mITT population had 185 subjects (N=90 for placebo and N=95 for pimavanserin).

[7] The unique subject ID's of the 14 patients who were excluded from mITT population were: ACP-103-020-038-104, ACP-103-020-039-103, ACP-103-020-301-108, ACP-103-020-303-122, ACP-103-020-318-104, ACP-103-020-320-104, and ACP-103-020-330-101.

Patient Disposition (Table 5)

The patient disposition and the number of patients who had efficacy assessment at Week 6 are provided in the following table.

Table 5: Patient Disposition for Primary Efficacy Assessment with Discontinuation Reason

Treatment group (# Randomized subjects)	**Placebo** (N=94)	**Pimavanserin** (N=105)	**Total** (N=199)
Excluded from mITT population	4	10	14
Sponsor mITT population	90	95	185
SAPS-PD observed at Week 6	86	87	173
SAPS-PD observation missing at Week 6	4	8	12
# Discontinued Patients	7	16	23
Reason for Discontinuation from Study (Sponsor mITT population)			
Adverse event	2	10	12
Voluntary withdrawal of consent	2	3	5
At discretion of ACADIA	2	2	4
Subject fails to comply with protocol requirements	0	1	1
Investigator's decision	1	0	1

Source: Table 14.1.2.1.1 of the CSR and Reviewer's analysis

3.2.5 Results and Conclusions

3.2.5.1 Primary Efficacy Endpoint (based on SAPS-PD)

Primary Analysis (Table 6)

The study efficacy conclusion was drawn from the primary analysis result. As listed in Table 6, pimavanserin's efficacy for the PDP (Hallucination and Delusion) was statistically demonstrated at the 0.05 significance level (2-sided test).

Table 6: Primary Efficacy Analysis Result (SAPS-PD Change Score from Baseline at Week 6)

Pimavanserin Difference from Placebo in LS Mean of Change from Baseline score (Week 6)				
LS Mean Estimate		Difference from Placebo in LS Mean Estimate (SE)	95% Confidence Interval	P value
Pimavanserin (SE)	Placebo (SE)			
-5.79 (0.66)	-2.73 (0.67)	-3.06 (0.94)	(-4.91, -1.20)	0.0014

Note: A negative change from baseline indicates an improvement. LS (Least Square) Mean estimates were obtained from an application of the pre-specified MMRM.

Table 7 provides visit-wise LS mean estimates, and the difference of pimavanserin from placebo in LS mean estimate and its 95% confidence interval with the nominal p value.

Table 7: Visit-wise Efficacy based on Change Score from Baseline of SAPS-PD (via Primary Analysis Model)

LS Mean Estimate by Treatment group (Change from baseline score at Weeks 2, 4 and 6)					
Treatment group	Visit	LS Mean Estimate (SE)	Difference from Placebo in LS Mean Estimate (SE)	95% Confidence Interval	P value
Pimavanserin	Week 2	-3.1 (0.60)	-0.21 (0.86)	(-1.9,-1.5)	0.8092
	Week 4	-5.0 (0.61)	-1.82 (0.87)	(-3.5, -0.1)	0.0369
	Week 6	-5.8 (0.66)	-3.06 (0.94)	(-4.9, -1.2)	0.0014
Placebo	Week 2	-2.9 (0.61)			
	Week 4	-3.2 (0.61)			
	Week 6	-2.7 (0.67)			

Note: LS (Least Square) Mean estimates were obtained from an application of the pre-specified MMRM. SE denotes standard error. Confidence interval and p-values were obtained without adjusting for multiplicity. Source: Reviewer's analysis

Sensitivity Analysis (Table 8)

The table below lists two of the sponsor's sensitivity analyses using the mITT population. The ANCOVA using an imputation of WOCF (worst observation carried forward) was an approach that is the least favorable to pimavanserin of the sponsor's pre-specified sensitivity analyses, a scenario that all missing observations were the worst of each patient's observed values. The result of ANCOVA (WOCF) does not suggest that the conclusion from the primary analysis may be questioned. The ANCOVA (LOCF) is equivalent to an ANCOVA using last observed values. The result of ANCOVA (LOCF) suggests that the dropouts only favored pimavanserin efficacy estimation by the amount of 0.15, a difference between -3.06 (primary analysis) and -2.91 (ANCOVA (LOCF)).

Table 8: SAPS-PD – Sensitivity Analyses versus Primary Analysis

	Method	LS Mean Change from Baseline at Week 6 (SE)		Difference from Placebo (pimavanserin – placebo)	95% Confidence Interval	p-value
		Pimavanserin	Placebo			
Primary Analysis	MMRM	-5.79 (0.66)	-2.73 (0.67)	-3.06 (0.94)	(-4.91, -1.20)	0.001
Sensitivity Analysis	ANCOVA (LOCF)	-5.56 (0.65)	-2.65 (0.67)	-2.91 (0.93)	(-4.76, -1.07)	0.002
	ANCOVA (WOCF)	-5.43 (0.65)	-2.65 (0.67)	-2.78 (0.94)	(-4.63, -0.93)	0.003

Note: LS (Least Square) Mean estimates were obtained from an application of the pre-specified MMRM. SE denotes standard error. Source: Table 13 of the CSR and Reviewer's analysis.

Observed Data used in Primary Efficacy Evaluation (Table 9)

The overall rate of dropouts for the efficacy assessment at Week 6 was around 6.5%, and there was not a large difference in the rate of missing observations between Pimavanserin and Placebo groups. The observed raw mean changes from baseline in SAPS-PD score show an improvement of 6.3 in SAPS-PD score (Pimavanserin) and 2.7 (Placebo). At Week 6, the observed raw mean difference between the two treatment groups was 3.6 points in favor of pimavanserin.

Table 9: SAPS-PD – Raw Mean Scores of Pimavanserin and Placebo Groups by Visit

	SAPS-PD Observed Visit-wise Raw Score of mITT Patients (N=185)					
Treatment group	Visit	N (% of Remaning Patients)	SAPS-PD total score Mean change score from baseline (SD)	SAPS-PD total score Mean observed score (SD)	SAPS-PD Delusion score Mean observed score (SD)	SAPS-PD Hallucination score Mean observed score (SD)
Pimavanserin	Baseline	95	-	15.9 (6.12)	4.8 (3.59)	11.1 (4.58)
	Week 2	94 (98.9)	-3.3 (6.13)	12.6 (7.45)	3.5 (3.59)	9.1 (5.29)
	Week 4	88 (92.6)	-5.3 (6.19)	10.8 (7.20)	3.0 (3.30)	7.8 (5.00)
	Week 6	87 (91.6)	-6.3 (5.88)	9.7 (7.11)	2.7 (3.02)	7.0 (5.14)
Placebo	Baseline	90	-	14.7 (5.55)	4.8 (3.82)	10.0 (3.80)
	Week 2	90 (100)	-2.7 (6.08)	12.1 (6.39)	3.6 (3.39)	8.5 (4.47)
	Week 4	89 (98.9)	-3.0 (6.41)	11.7 (6.11)	3.5 (3.26)	8.2 (4.50)
	Week 6	86 (95.6)	-2.7 (7.03)	11.8 (7.17)	3.6 (3.95)	8.3 (4.43)

SD denotes standard deviation.
Source: Reviewer's analysis

Table 10: Sponsor Supportive Analysis Results (SAPS-H+D)

Pimavanserin vs Placebo at Week 6 Change from baseline score in Supportive endpoints					
Efficacy Measure	LS Mean Estimate (SE)		Difference from Placebo in LS Mean Estimate (SE)	95% Confidence Interval	P value
	Pimavanserin	Placebo			
SAPS-H+D	-6.51 (0.72)	-3.14 (0.73)	-3.37 (1.03)	(-5.40, -1.35)	0.0012
SAPS-H	-4.18 (0.49)	-2.10 (0.49)	-2.08 (0.70)	(-3.46, -0.71)	0.0032
SAPS-D	-2.28 (0.38)	-1.12 (0.38)	-1.16 (0.54)	(-2.22, -0.10)	0.0325
GSAPS-H+D	-1.95 (0.26)	-1.02 (0.26)	-0.93 (0.37)	(-1.65, -0.21)	0.0117
GSAPS-H	-1.16 (0.15)	-0.50 (0.15)	-0.66 (0.21)	(-1.08, -0.25)	0.0020
GSAPS-D	-0.80 (0.15)	-0.53 (0.15)	-0.27 (0.21)	(-0.68, 0.14)	0.1890

Note: The analysis method was the same as the primary efficacy analysis (MMRM). SE denotes standard error. GSAPS denotes Combined SAPS-Hallucinations and Delusions Global Rating of Severity.
Source: Tables 14.2.3.5.1, 14.2.3.7.1, 14.2.3.8.1, 14.2.3.6.1, 14.2.3.9.1 and 14.2.3.10.1 of the CSR

Supportive Analysis (Table 10)

The sponsor planned and performed supportive analyses using SAPS-H+D (20-item SAPS score for hallucination and delusion). See Section 3.2.3.3 of this review. Overall, these analysis results are supportive for the primary efficacy analysis conclusion. It is noted that the numerical results for the delusion domain appear to be slightly less supportive compared to those for the hallucination domain.

3.2.5.2 Key Secondary Endpoint (based on UPDRS II+III)

Primary Analysis and Non-inferiority Test (Table 11)

The key secondary endpoint was the mean change from baseline to Week 6 in the combined score of the UPDRS Part II (activities of daily living) and Part III (motor examination) using the ANCOVA analysis (OC) for the mITT analysis set. Essentially, this analysis is equivalent to an analysis using only completersbecause there was only one post-baseline assessment (at week 6). It is noted that there were only a few missing observations at the endpoint visit in total (Table 12).

Since the upper limit (2.72) of the 2-sided 95% CI for the treatment difference did not exceed the pre-specified non-inferiority margin of 5 units, non-inferiority of pimavanserin 40 mg compared to placebo was established (Table 11). The sponsor concluded that the result suggests that there was no clinically meaningful difference in the activities of daily living and the motor examinations of the subjects taking pimavanserin 40 mg when compared with the subjects taking placebo.

Table 11: Primary Analysis Results of Key Secondary Endpoint: Parkinson's Disease Status (UPDRS II+III)

Key Secondary Endpoint: Pimavanserin vs Placebo at Week 6 Change score from baseline in UPDRS II+III Combined score				
ANCOVA (OC) LSM Estimate (SE)		Difference from Placebo in LSM Estimate (SE)	95% Confidence Interval	P value
Pimavanserin [N=92]	Placebo [N=88]			
-1.40 (0.86)	-1.69 (0.88)	0.29 (1.23)	(-2.14, 2.72)	0.8140

Note: A negative change from baseline indicates an improvement. The analysis result is based on ANCOVA (OC) model with treatment group as a factor and baseline score as a covariate. OC denotes Observed Cases. SE denotes standard error. N denotes the number of patients who had a baseline score and the endpoint score at Week 6.
Source: Reviewer's analysis

Table 12: Raw Mean Change Score from Baseline at Week 6 (UPDRS II+III)

Mean change from baseline at Week 6 (SD)	Pimavanserin (N=95)		Placebo (N=90)	
UPDRS II+III combined score	-1.36 (8.16)	n=92	-1.73 (8.76)	n=88
UPDRS II (Activities in Daily Living) score	-0.52 (3.32)	n=92	-0.88 (3.33)	n=88
UPDRS III (Motor Examination) score	-0.80 (7.12)	n=93	-0.86 (7.09)	n=88

SD denotes standard deviation. N denotes the numbers of mITT patients, and n those of patients who had the endpoint score at Week 6.
Source: Tables 27 and 28 of the CSR and Reviewer's analysis

Correlation between SAPS-PD and UPDRS II+III (Table 13, Figure 3)

Out of 185 mITT patients, 175 patients had an assessment at Week 6 in both SAPS-PD and UPDRS II+III.

In Figure 3, a graph of scatterplots between the changes from baseline at Week 6 of the SAPS-PD and UPDSR II and III combined scores is provided for these 175 patients. It seems that there is no tendency suggesting that patients improve on psychosis but worsen on Parkinson's Disease status.

Mean changes from baseline at Week 6 of both scores are listed for these 175 patients in Table 13.

Table 13: Mean Changes from Baseline at Week 6 of SAPS-PD total score and UPDRS II+III combined score (N=175)

Mean change from baseline at Week 6 (SD)	Pimavanserin (N=95)		Placebo (N=90)	
SAPS-PD total score	-6.14 (6.11)	n=88	-2.70 (7.00)	n=87
UPDRS II+III combined score	-1.64 (7.89)		-1.76 (8.80)	

Note: n denotes number of patients who had both scores at Week 6; SD denotes standard deviation.
Source: Reviewer's analysis

3.2.5.3 Other Secondary Efficacy Endpoints (Table 14, Table 15)

The MMRM analysis results of secondary efficacy endpoints based on CGI-S and CGI-I are listed in Table 14 and Table 15. Both results are supportive for pimavanserin's efficacy.

Table 14: Secondary Efficacy (CGI-S)

Pimavanserin Difference from Placebo in Mean change from baseline score of CGI-S at Weeks 2, 4 and 6			
Visit	LS Mean Estimate (SE)	95% Confidence Interval	P value
Week 2	-0.02 (0.15)	(-0.32,-0.27)	0.8730
Week 4	-0.41 (0.18)	(-0.76, -0.06)	0.0224
Week 6	-0.58 (0.17)	(-0.92, -0.25)	0.0007

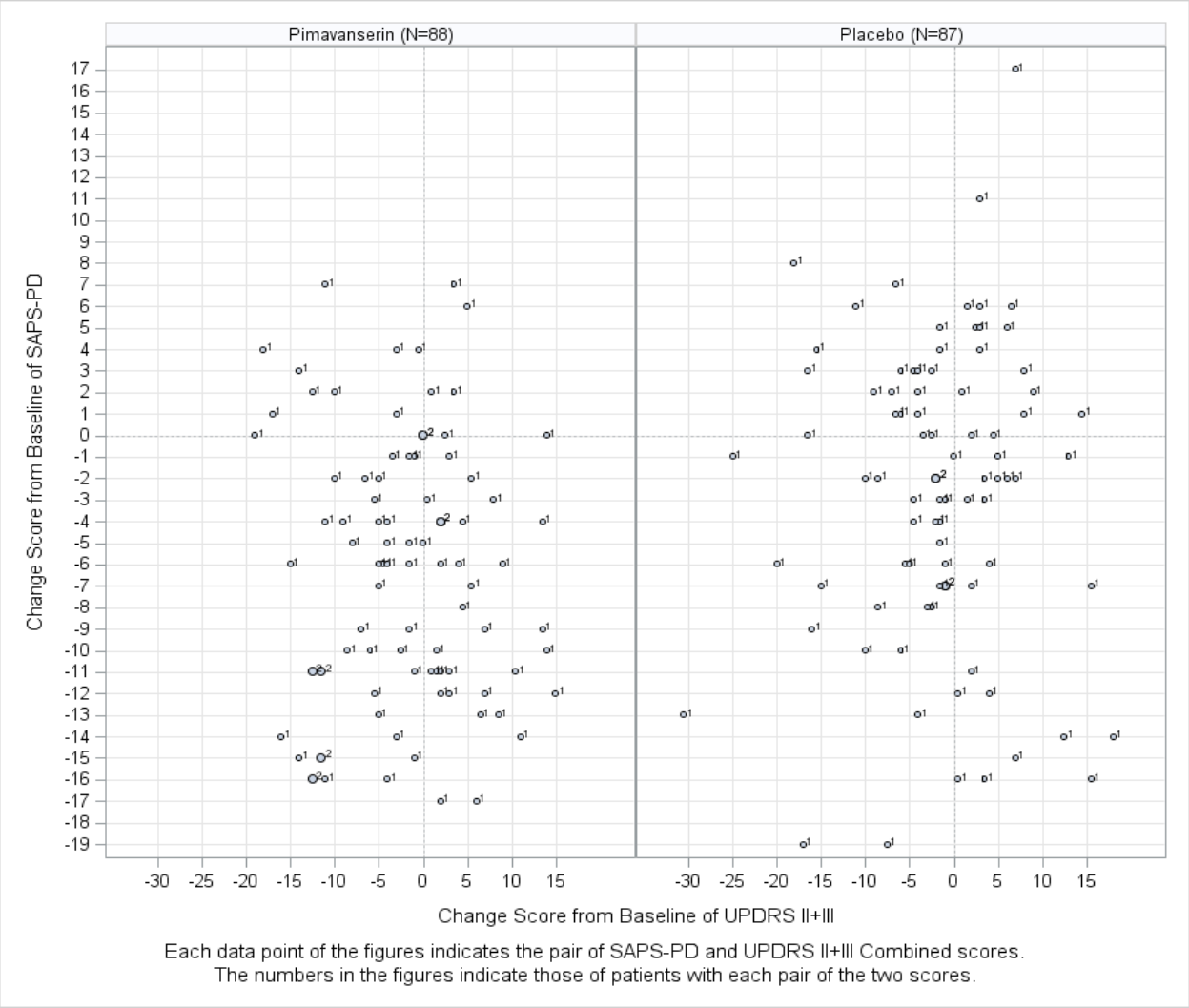
SE denotes standard error.
Source: Table 14.2.2.2.1 of the CSR and Reviewer's analysis

Table 15: Secondary Efficacy (CGI-I)

Pimavanserin Difference from Placebo in Mean score of CGI-I at Weeks 2, 4 and 6			
Visit	LS Mean Estimate (SE)	95% Confidence Interval	P value
Week 2	-0.01 (0.18)	(-0.36,-0.33)	0.9363
Week 4	-0.50 (0.19)	(-0.88, -0.12)	0.0098
Week 6	-0.67 (0.20)	(-1.06, -0.27)	0.0011

SE denotes standard error.
Source: Table 14.2.2.3.1 of the CSR and Reviewer's analysis

Figure 3: Scatterplots of SAPS-PD and UPDRS II+III Change Scores from Baseline to Week 6 by Treatment Group (N=175)



Source: Reviewer's analysis

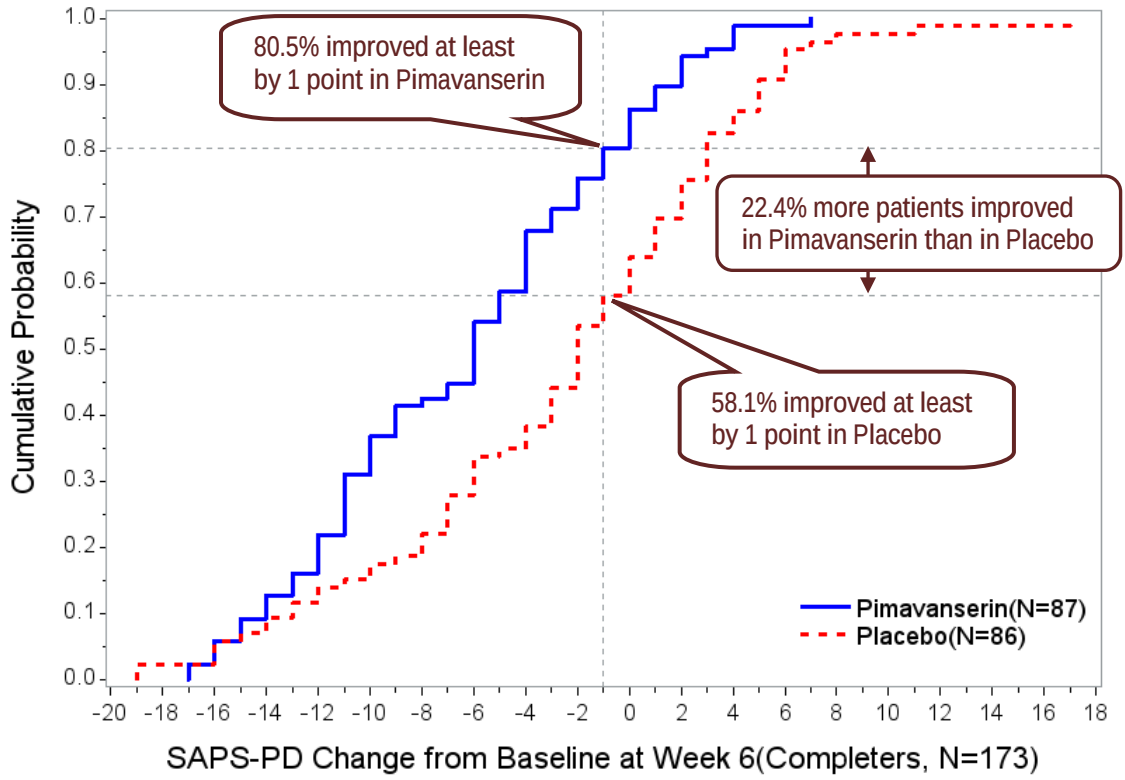
3.3 Exploratory Analysis

The exploratory analyses provided in this section are based on the mITT patients (N=185). Analyses of this section are all *post hoc*, and cannot be interpreted as conclusive or generalizable. The statistical results (estimates, confidence intervals, p values) shown in this section should not be used to draw any general conclusion. This section only provides descriptive information about the study efficacy data.

3.3.1 Empirical Cumulative Distribution of SAPS-PD Change from Baseline Score (Figure 4)

The empirical cumulative distribution plots provide visualized efficacy of pimavanserin in comparison with placebo. Figure 4 shows the benefit of pimavanserin as a treatment of PDP. A patient improves when the change from baseline score is less than zero. As shown in the figure, 80.5% of pimavanserin completers and 58.1% of placebo completers improved. A cumulative probability of pimavanserin was around 16% higher on average than that of placebo, and the placebo response seems to have been considerably high.

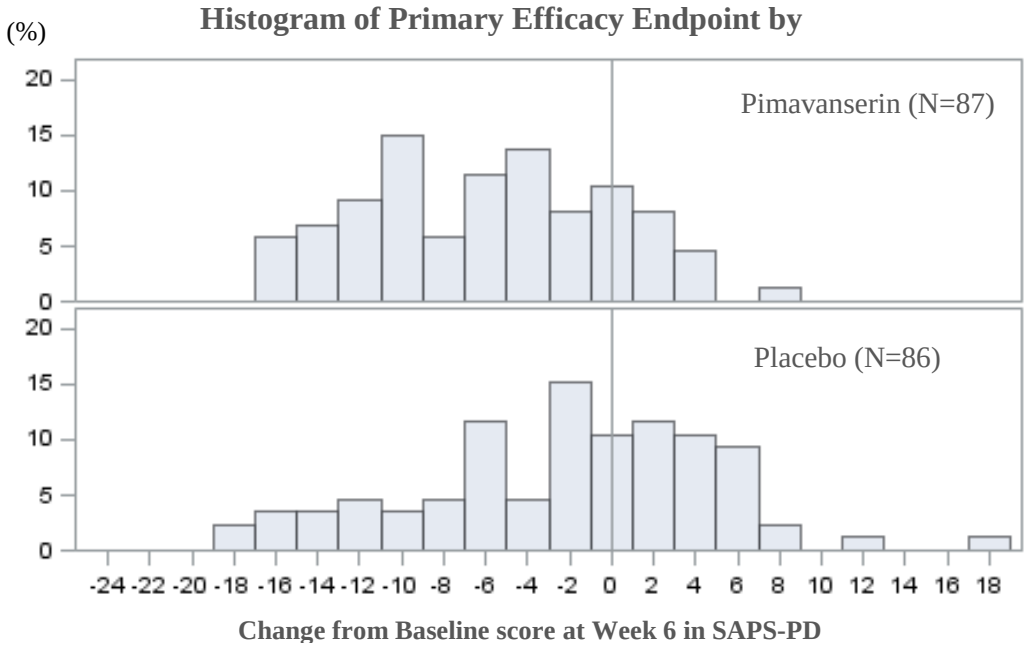
Figure 4: Empirical Cumulative Distribution Function Plots by Treatment at Week 6 (Observed Change Scores from Baseline in SAPS-PD) (Completers, N=173)



Note: A patient improves if the change from baseline score to Week 6 is less than 0.
Source: Reviewer's analysis

Figure 5 shows the histogram of primary endpoint scores (change from baseline to Week 6 in SAPS-PD) for 173 completers. The distributional ranges are similar for both treatment groups, but the histogram of pimavanserin group does not look skewed, while that of placebo group is skewed left, which is an indication that pimavanserin is more efficacious than placebo.

Figure 5: Empirical Distribution of Change from Baseline to Week 6 in SAPS-PD (for 173 Completers)



Source: Reviewer's analysis

3.3.2 Scatterplots of Individual and Global Item Sums of SAPS-PD at Week 6 (Figure 5)

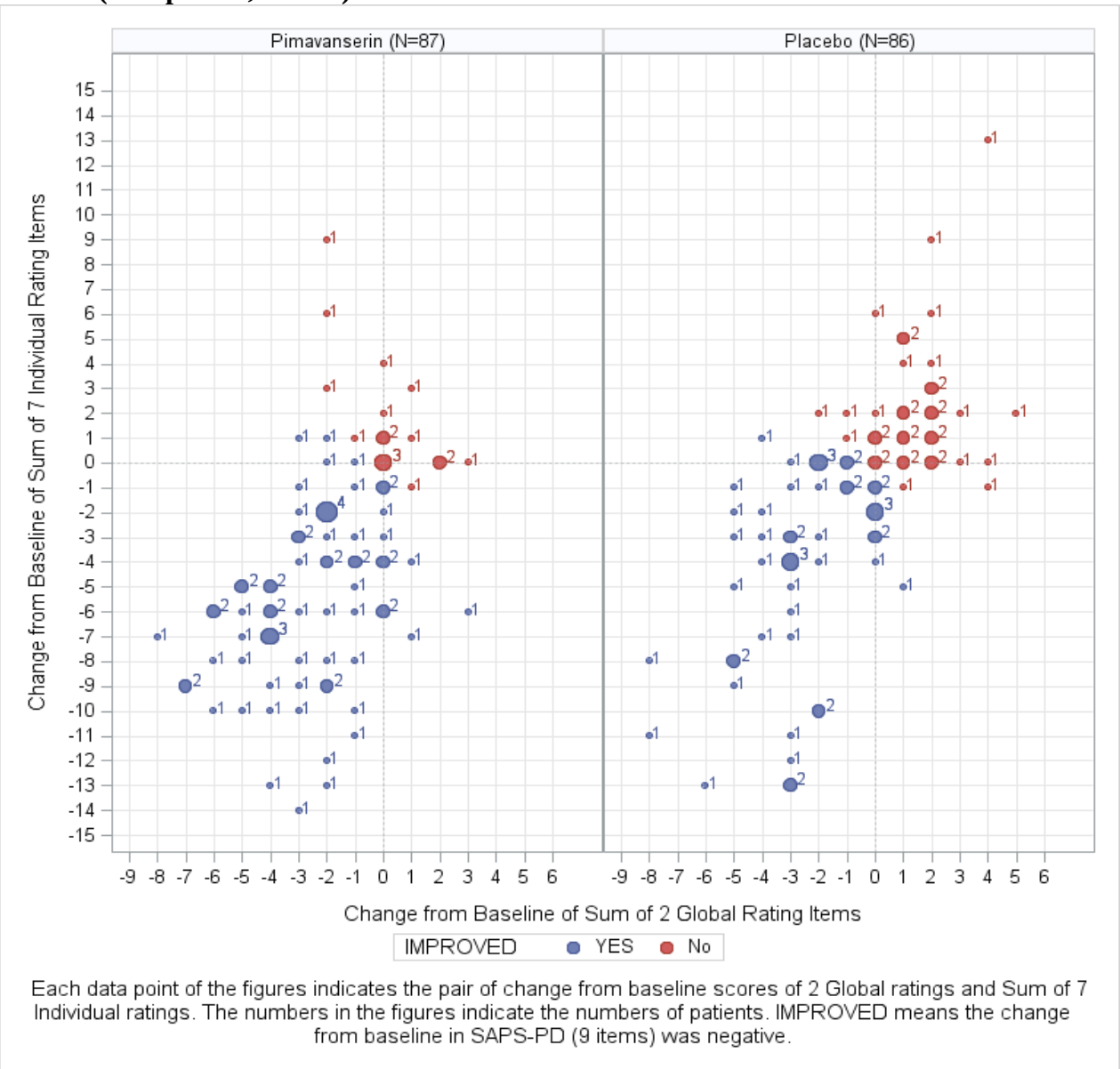
SAPS-PD is a new scale for evaluating psychosis (hallucination and delusion) that is adopted from SAPS-H+D. SASP-PD consists of 9 items: 5 Hallucination domain scores (4 individual items and one global item) and 4 Delusion domain scores (3 individual items and one global item)). In the original SAPS[8] (consisting of 4 domains of Hallucination, Delusion, Bizarre Behavior and Positive Formal Thought Disorder), it is said that the global item scores additionally cover other positive symptoms than the individual items of the respective domain capture. Therefore, the associative relationship between the sum of 7 individual items and the sum of 2 global items of SAPS-PD may not show a strong associative relationship.

Figure 6 shows scatterplots of the sum of 7 individual item scores versus the sum of 2 global item scores for the 173 completers. Each circle denotes a configuration of the changes from baseline in the paired sums. The number next to each circle denotes the number of patients with the given configuration; a larger circle indicates more patients. In both treatment groups, there seems to be positive correlation, but the global item scores were not highly correlated with the sum of the 7 individual item scores. Evidence of pimavanserin efficacy is visualized in the plots.

[8] Scale for the Assessment of Positive Symptoms (SAPS) by Nancy C. Andreasen, 1984

"Improved patients (those with negative changes from baseline in the SAPS-PD (the sum of all 9 items)" are colored in blue, and "non-improved patients" in red. It appears that pimavanserin had a better efficacy than placebo both in *improved patients* and *non-improved patients*. Clearly, a much higher proportion of patients improved in pimavanserin than in placebo, but some placebo patients had a great improvement that is as good as pimavanserin patients who improved greatly.

Figure 6: Scatterplots of Individual versus Global Sums of SAPS-PD by Treatment at Week 6 (Completers, N=173)



Each data point of the figures indicates the pair of change from baseline scores of 2 Global ratings and Sum of 7 Individual ratings. The numbers in the figures indicate the numbers of patients. IMPROVED means the change from baseline in SAPS-PD (9 items) was negative.

Source: Reviewer's analysis

3.3.3 Exploratory Efficacy Analysis of Sum of 7 Individual Items of SAPS-PD (Table 16)

The SAP-PD (9-item SAPS) consists of 7 individual items measuring hallucination and delusion symptoms and 2 (Hallucination and Delusion) global items. This reviewer analyzed the sum of 7 individual items of SAPS-PD, using the same analysis method as in the primary analysis; the difference from placebo in the LS mean estimate was -2.01 with a p value of 0.0022. The difference from placebo in the LS mean estimate for the primary endpoint (sum of the 9 items) was -3.06.

Table 16: Pimavanserin Efficacy for Sum of 7 individual item scores of SAPS-PD

PIM 40 mg vs Placebo at Week 6/Change from baseline score in Supportive endpoints				
LSM Estimate (SE)		Difference from Placebo in LSM Estimate (SE)	95% Confidence Interval	P value
Pimavanserin	Placebo			
-3.80 (0.45)	-1.79 (0.46)	-2.01 (0.64)	(-3.28, -0.74)	0.0022

Note: The analysis method was the same as the primary efficacy analysis (MMRM). SE denotes standard error.
Source: Reviewer's analysis

A similar analysis for the sum of the two global item scores also shows a statistical significance with a p value of 0.0117. The difference from placebo in LS mean estimate was -0.93. This analysis is provided by the sponsor in their supportive analyses. The result is displayed in Table 10 of this review (under the title of GSAPS H+D).

3.3.4 Exploratory Efficacy Analysis of SAPS-PD Hallucination and Delusion Domains

Table 17 and Table 18 display the MMRM analysis result for each of the domains (Delusion and Hallucination) of SAPS-PD for the primary analysis set (N=185).

The scatter plots shown in Figure 7 and Figure 8 provide a visualized positive correlation for 173 completers between the sum of individual scores and the global score for each domain. By construct, the global score measures patient overall delusion or hallucination severity that may not be covered by the individual scores. A visual inspection of the plots may give an impression that in delusion and hallucination, respectively, the pimavanserin efficacy came more from patients who did not improve (indicated by circles in red) rather than from those who improved (indicated by circles in blue). Interestingly, many patients had no improvement or worsening outcome at Week 6 (a score of zero in change from baseline in SAPS-PD) in each of the domains, which was not seen in the figure for the SAPS-PD total score (Figure 6).

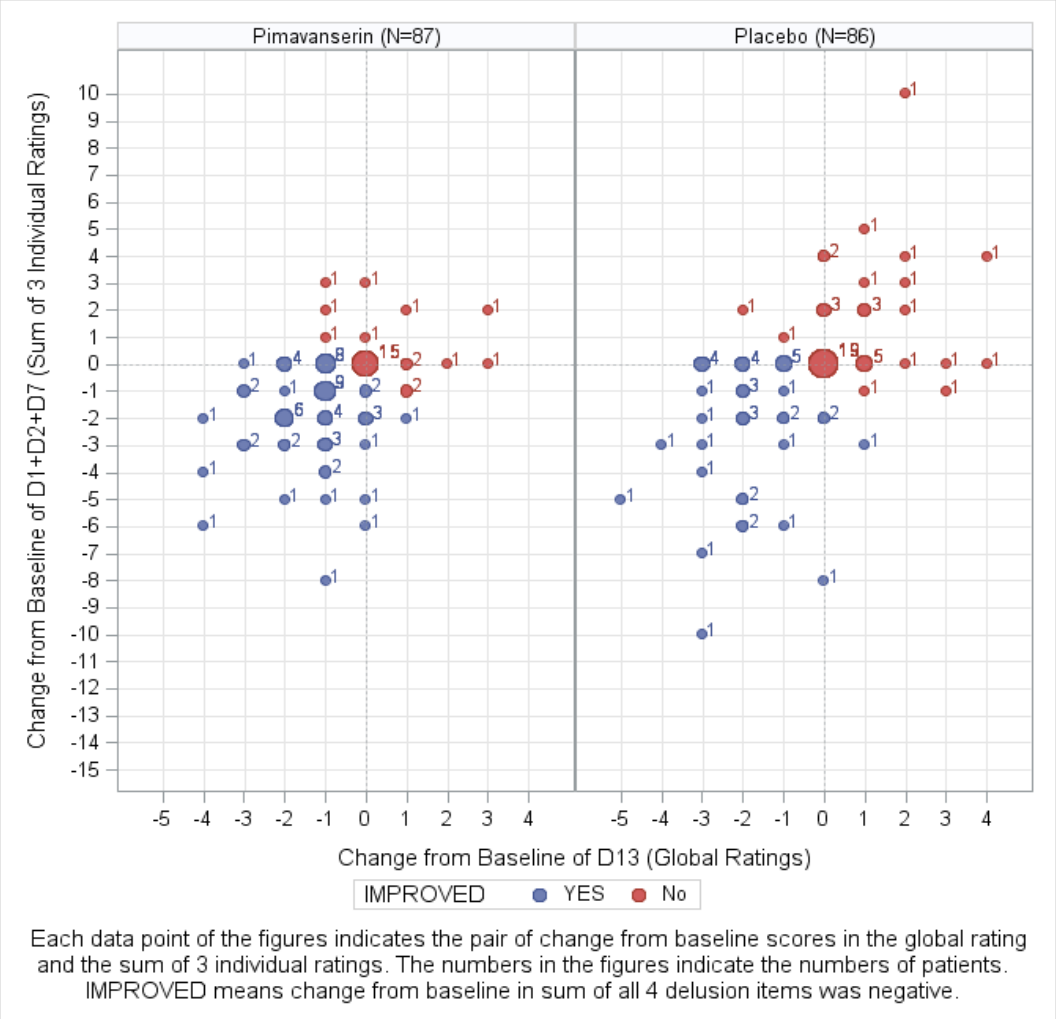
3.3.4.1 Delusion Domain (Sum of 4 Delusion items of SAPS-PD) (Table 17, Figure 7)

Table 17: Delusion Domain of SAPS-PD – Exploratory Efficacy Analysis (mITT Population, N=185)

Pimavanserin vs Placebo at Week 6 (Change from baseline score of SAPS-PD Delusion Domain score)				
LS Mean Estimate (SE)		Difference from Placebo in LS Mean Estimate (SE)	95% Confidence Interval	P value
Pimavanserin	Placebo			
-1.95 (0.32)	-1.01 (0.32)	-0.94 (0.45)	(-1.83, -0.04)	0.0403

Note: The analysis method was the same as the primary efficacy analysis (MMRM). SE denotes standard error.
Source: Reviewer's analysis

Figure 7: Scatterplots of Individual Item Score Sum versus Global Score for Delusion Domain of SAPS-PD by Treatment at Week 6 (Completers, N=173)



Source: Reviewer's analysis

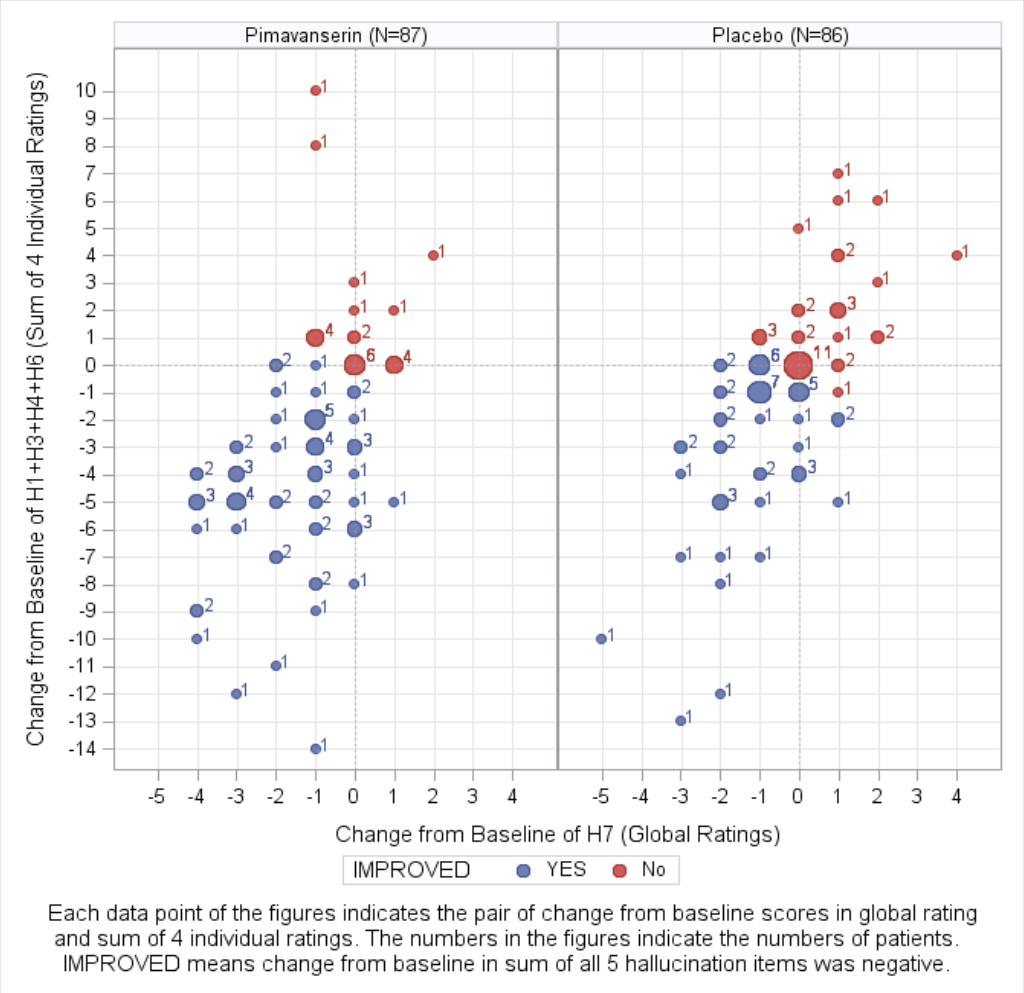
3.3.4.2 Hallucination Domain (Sum of 5 items of SAPS-PD) (Table 18, Figure 8)

Table 18: Hallucination Domain of SAPS-PD – Exploratory Efficacy Analysis (mITT Population, N=185)

PIM 40 mg vs Placebo at Week 6 (Change from baseline score of SAPS-PD Hallucination Domain score)				
LS Mean Estimate (SE)		Difference from Placebo in LS Mean Estimate (SE)	95% Confidence Interval	P value
Pimavanserin	Placebo			
-3.81 (0.46)	-1.80 (0.46)	-2.01 (0.65)	(-3.30, -0.72)	0.0024

Note: The analysis method was the same as the primary efficacy analysis (MMRM). SE denotes standard error.
Source: Reviewer's analysis

Figure 8: Scatterplots of Individual Item Score Sum versus Global Score for Hallucination Domain of SAPS-PD by Treatment at Week 6 (Completers, N=173)



Source: Reviewer's analysis

3.4 Evaluation of Safety

Safety evaluation was not conducted in this review.

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race, Age, and Geographic Region (Table 19)

The sponsor included subgroup summary statistics for gender, race (white or non-white), age group (sponsor specified thresholds of 65 and 75 years of age). The sponsor concluded that their subgroup analyses show a consistent trend for superiority of pimavanserin over placebo regardless of subject age, gender (page 63 of the CSR). This reviewer has no objection to the sponsor's conclusion. As this study was conducted in US and Canada with only 5 patients from Canada, a subgroup analysis by region/country is not performed.

For each treatment group, the summary statistics of observed scores and change from baseline scores of SAPS-PD were included in the CSR (Tables 14.2.3.14 - 14.2.3.16). For a reference purpose provided below is a table of selected summary statistics of the sponsor's subgroups (gender, race and age group) from the sponsor's CSR tables.

Table 19: Sponsor Subgroup Analysis (Gender, Race, Country and Age Group)

Observed (Raw) data Primary Endpoint (SAPS-PD)		Mean Change from Baseline at Week 6 (SD)				
		Pimavanserin	#Subjects	Placebo	#Subjects	Total Number of Subjects at Week 6
Primary Analysis Set mITT population (N=185)		-6.3 (5.88)	87	-2.7 (7.03)	86	173
Gender	Male	-7.3 (5.54)	56	-3.0 (6.82)	50	106
	Female	-4.7 (6.20)	31	-2.2 (7.38)	36	67
Race	White	-6.0 (5.81)	82	-2.3 (6.83)	81	163
	Non-white	-11.4 (5.18)	5	-7.8 (9.04)	5	10
Age Group	< 65 years of age	-5.0 (4.60)	10	-5.4 (5.14)	11	21
	≥ 65 and < 75 years of age	-6.5 (6.14)	47	-2.3 (6.77)	47	94
	> 75 years of age	-6.6 (5.96)	30	-5.4 (8.00)	28	58

SD denotes standard deviation.
Source: Tables 14.2.3.14 - 14.2.3.16 of the CSR

4.2 Other Special/Subgroup Populations

4.2.1 MMSE (Mini Mental State Exam)

The sponsor conducted a subgroup analysis based on MMSE baseline score. The result is provided in the following table.

Table 20: Sponsor Subgroup Analysis by MMSE baseline score

Observed (Raw) data Primary Endpoint (SAPS-PD)		Mean Change from Baseline at Week 6 (SD)				
		Pimavanserin		Placebo		Total Number of Subjects at Week 6
			#Subjects		#Subjects	
Primary Analysis Set mITT population (N=185)		-6.3 (5.88)	87	-2.7 (7.03)	86	173
MMSE	≥25	-6.0 (5.78)	60	-3.3 (7.38)	67	127
	<25	-7.1 (6.14)	27	-0.5 (5.18)	19	46

Source: Table 14.2.3.17 of the CSR

4.2.2 Subgroup by Delusion Baseline Severity (Figure 9, Table 21, Table 22)

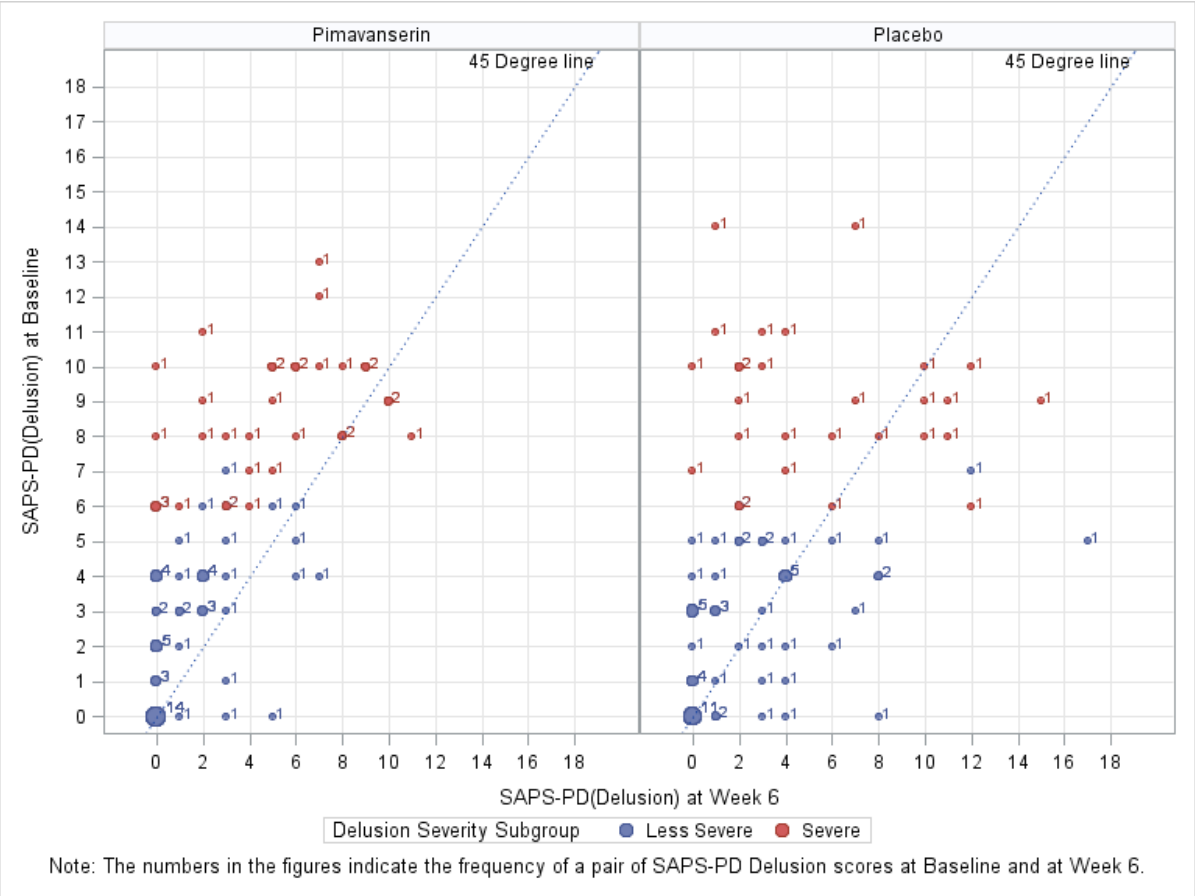
As seen in Figure 2, at baseline, 114 patients had a score in SAPS-PD Delusion in the lower range (0 – 5 points), which may suggest that those patients were enrolled in the study for hallucination but not for delusion.

This reviewer created a subgroup based on delusion severity at baseline using SAPS-PD Delusion Domain score. The criteria this reviewer used to categorize patients with a delusion baseline status of 'severe' or 'less severe' are as follows: Those who had at least 3 points both in the delusion global score and the sum of delusion non-global scores are in a status of "severe in delusion." A patient who did not have at least 3 points both in the global score and the sum of non-global scores in SAPS-PD Delusion Domain at baseline may be categorized as "less severe in delusion."

Figure 9 visualizes an improvement from baseline at Week 6 in SAPS-PD Delusion Domain score. It plots individual pairs of baseline and Week 6 scores of SAPS-PD Delusion Domain. Patients who are located above the 45 degree line drawn in the figure were those who improved at Week 6 endpoint (Baseline score > Week 6 score). Pairs plotted in red indicate patients who had relatively severer SAPS-PD Delusion score at baseline, and those in blue indicate patients who had less severe SAPS-PD Delusion score at baseline.

Table 21 shows mean baseline score and raw mean change score from baseline of SAPS-PD Delusion Domain score (ranging from 0 – 20 points with 4 item scores) by treatment group for overall patients (mITT population), and for severity subgroups. It appears that patients with less severe delusion at baseline contributed somewhat more to the treatment effect with respect to this domain.

Figure 9: Mapping individual patient improvements on SAPS-PD Delusion Domain Score from Baseline to Week 6 by Delusion Severity (Completers, N=173)



Source: Reviewer's analysis

Table 21: Mean Baseline Sore and Raw Mean Change Score from Baseline of SAPS-PD Delusion Domain (Completers, N=173)

SAPS-PD Delusion	Pimavanserin			Placebo			Pimavanserin Mean Difference from Placebo in Change from Baseline
	N	Mean (SD)		N	Mean (SD)		
		Baseline	Change from Baseline		Baseline	Change from Baseline	
Overall	87	4.71 (3.60)	-2.01 (2.70)	86	4.56 (3.70)	-1.01 (4.09)	-1.00
Severe	31	8.71 (1.79)	-3.65 (3.01)	27	9.11 (2.04)	-3.37 (5.11)	-0.28
Less Severe	56	2.50 (2.11)	-1.11 (2.02)	59	2.47 (2.01)	0.07 (3.01)	-1.18
Removing patients who had a score of zero both at baseline and Week 6	42	3.33 (1.78)	-1.48 (2.22)	48	3.04 (1.80)	0.08 (3.34)	-1.56

Source: Reviewer's analysis

Table 22 shows the MMRM analysis result for the delusion subgroup analysis. This analysis is *post hoc* exploratory, and as such has a limitation in the generalizability of the result. The analysis may be underpowered and the inferential result should be carefully interpreted.

However, the result seems to suggest that the subgroup of patients of less severe in delusion may have had more of a share in efficacy evidence shown in the overall delusion analysis based on the combined subgroups of severe and less severe patients. The magnitude of difference may not be clinically relevant for the indication of hallucination and delusion.

Table 22: Exploratory MMRM Analysis - Change Score from Baseline of SAPS-PD Delusion Domain (mITT Population, N=185)

Pimavanserin vs Placebo at Week 6 (Change from baseline score of SAPS-PD Delusion total score)				
Baseline Severity in SAPS-PD	LS Mean Estimate (SE)		Difference from Placebo in LS Mean Estimate (SE)	95% Confidence Interval
	Pimavanserin	Placebo		
Overall	-1.95 (0.32)	-1.01 (0.32)	-0.94 (0.45)	(-1.83, -0.04)
Severe	-3.83 (0.68)	-3.07 (0.72)	-0.76 (0.99)	(-2.74,1.22)
Less Severe	-1.00 (0.33)	0.10 (0.32)	-1.10 (0.46)	(-2.02,-0.19)

Note: The analysis method was MMRM with the same model as in the primary efficacy. SE denotes standard error.
Source: Reviewer's analysis

Note on Hallucination Severity Subgroup

A subgroup based on hallucination severity at baseline may not be meaningful. As shown in Figure 2, there were only a small number of patients in the mITT analysis set who had a considerably low (less than 4, say) baseline score of SAPS-PD Hallucination Domain.

5 SUMMARY AND CONCLUSIONS

The single pivotal study (Study ACP-103-020) has established statistical evidence that pimavanserin 40 mg is efficacious as a treatment of Parkinson's Disease Psychosis (PDP) and does not worsen Parkinson's Disease status. No major issue that may affect the main statistical conclusions was found.

The primary efficacy endpoint was the change from baseline to Week 6 in score of a novel instrument SAPS-PD (Scale for the Assessment of Positive Symptoms for Parkinson's Disease). The SAPS-PD assesses Parkinson's Disease Psychosis (PDP) for Hallucination and Delusion symptoms. Based on SAPS-PD, pimavanserin 40 mg has been shown to be more efficacious than placebo for an acute treatment of PDP.

The key secondary endpoint was the change from baseline to Week 6 in UPDRS Parts II and III Combined score, the respective part of which assesses activities in daily life and motor functions of Parkinson's Disease patients. A non-inferiority of pimavanserin to placebo has been concluded based on a non-inferiority test with a non-inferiority margin of 5 points. To be specific, pimavanserin ruled out >2.72 points worse than placebo based on the 95% confidence interval of the treatment effect relative to placebo.

A Psychopharmacologic Drugs Advisory Committee (PDAC) meeting will be held on March 29, 2016. One of the questions addressed to the PDAC may likely be about whether the treatment effect observed in this study, 3 points in SAPS-PD as an observed effect (a difference of pimavanserin from placebo in least square mean estimate), is clinically meaningful. This review may provide information on clinical effectiveness of the candidate new treatment.

The following findings from exploratory and subgroup analyses, and other analyses, may help understand the clinical meaningfulness of the observed treatment effect.

1. Many study patients seem to have had relatively mild delusion symptoms at baseline. In addition, a number of patients had a score of zero for delusion in SAPS-PD (33 patients, 16.6% of 199 randomized patients) and in NPI-H+D (58 patients, 31.4% if 185 mITT patients). (See Figure 1 and Figure 2.)
2. Exploratory SAPS-PD subscale analysis: This reviewer noted that the maximum total score for the Hallucination domain is larger than that for the Delusion domain (25 vs 20). Despite that, when each subscale (Hallucination and Delusion) is evaluated on the mITT patients, the treatment effect for primary efficacy seems to have a relatively larger share in Hallucination than in Delusion (-2.01 vs -0.94 in Difference from Placebo in LS mean estimates. See Table 17 and Table 18). However, the clinical relevancy may be discussed by the clinical review team.
3. Exploratory analysis for severity subgroup (subgroup based on baseline SAPS-PD subscale scores): A subgroup analysis based on baseline severity in delusion may suggest that:

a. Evidence of efficacy for delusion seems to have been dependent on improvements of patients who were less severe for delusion at baseline (See Table 21 and Table 22). However, the observed differences may not be clinically relevant.
b. Patients of both treatment groups who were severe in delusion at baseline seem to have had a large improvement on average in SAPS-PD Delusion Domain score, but the small difference between pimavanserin and placebo groups appears to be due to placebo response (See Table 22).

APPENDICES

A1 NPI (A and B)

A Delusions:

If the screening question is confirmed, determine the frequency and severity of the delusions.
Frequency:
1. Occasionally – less than once per week.
2. Often – about once per week.
3. Frequently – several times per week but less than every day.
4. Very frequently – once or more per day.
Severity:
1. Mild – delusions present but seem harmless and produce little distress in the patient.
2. Moderate – delusions are distressing and disruptive.
3. Marked – delusions are very disruptive and are a major source of behavioral disruption. [If PRN medications are prescribed, their use signals that the delusions are of marked severity.]

B Hallucinations:

If the screening question is confirmed, determine the frequency and severity of the hallucinations.
Frequency:
1. Occasionally – less than once per week.
2. Often – about once per week.
3. Frequently – several times per week but less than every day.
4. Very frequently – once or more per day.
Severity:
1. Mild – hallucinations are present but harmless and cause little distress for the patient.
2. Moderate – hallucinations are distressing and are disruptive to the patient.
3. Marked – hallucinations are very disruptive and are a major source of behavioral disturbance. PRN medications may be required to control them.

A2 SAPS-PD

The SAPS-PD is a 9 item instrument derived from the 20 item SAPS. The SAPS was developed as a clinician reported outcome through a semi-structure interview with a patient. The entire SAP was administered to each patient and the 9 items to create the SAPS-PD was pulled out to form a score. The 9 items are:

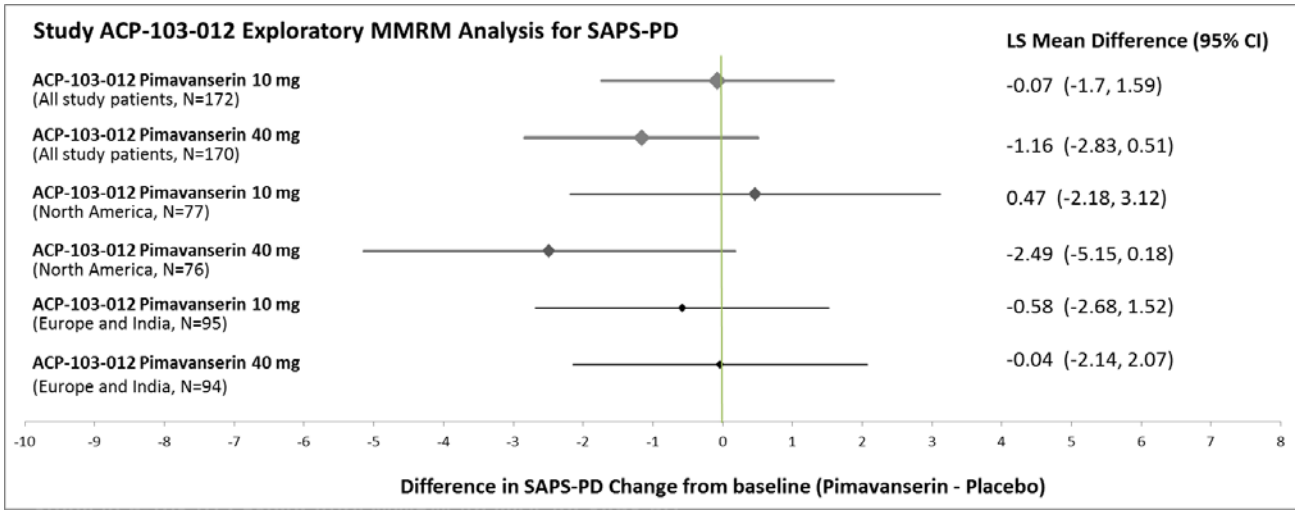
- H1 Auditory Hallucinations
- H3 Voices Conversing
- H4 Somatic or Tactile Hallucinations
- H6 Visual Hallucinations
- H7 Global Rating of Severity of Hallucinations
- D1 Persecutory Delusions
- D2 Delusions of Jealousy
- D7 Ideas and Delusions of Reference
- D13 Global Rating of Severity of Delusions

The range of scores for the SAPS-PD is 0-45 with a high score representing higher severity of psychosis. The responses were on a 0 through 5 NRS, with 0=none; 1=questionable, 2=mild, 3=moderate, 4=marked, and 5=severe. The SAPS was given to patients at baseline, Week 2, Week 4 and Week 6.

A3 Study ACP-103-012

Study ACP-103-012 investigated two pimavanserin doses: 10 mg and 40 mg. The following figure provides forest plots of pimavanserin's estimated difference from placebo (and the 95% confidence interval) in SAPS-PD change score from baseline obtained from the exploratory MMRM analysis (observed data only). The plots include results from subgroup analyses based on North America (US and Canada) and Outside of North America (Europe and India).

Figure 10: Study ACP-103-012 MMRM analysis results for SAPS-PD Change from Baseline at Week 6



Note: MMRM analysis was based on Observed data only (no imputation of missing data).

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

EIJI ISHIDA
02/22/2016

PEILING YANG
02/22/2016

HSIEN MING J HUNG
02/22/2016

CLINICAL OUTCOME ASSESSMENT CONSULT REVIEW

CLINICAL OUTCOME ASSESSMENT (COA) TRACKING NUMBER	AT 2015-180
IND/NDA/BLA NUMBER	NDA 207318
LETTER DATE/SUBMISSION NUMBER	9/1/2015
PDUFA GOAL DATE	5/1/2016
DATE OF CONSULT REQUEST	10/29/2015
REVIEW DIVISION	DPP
MEDICAL REVIEWER	Paul Andreason
REVIEW DIVISION PM	Brendon Muoio
PRIMARY COA REVIEWER	Michelle Campbell
SECONDARY COA REVIEWER	Wen-Hung Chen
ASSOCIATE DIRECTOR, COA STAFF (ACTING)	Elektra Papadopoulos
REVIEW COMPLETION DATE	2/5/2016
ESTABLISHED NAME	Pimavanserin
TRADE NAME	Nuplazid
SPONSOR/APPLICANT	ACADIA Pharmaceuticals Inc.
CLINICAL OUTCOME ASSESSMENT TYPE	Clinician Reported Outcome (ClinRO)
ENDPOINT(S) CONCEPT(S)	Decrease in Psychosis Symptoms
MEASURE(S)	Parkinson's Disease adapted Scale for the Assessment of Positive Symptoms (SAPS-PD)
INDICATION	Treatment of Psychosis associated with Parkinson's Disease
INTENDED POPULATION(S)	Adults aged 40 years or older with Psychosis associated with Parkinson's Disease

A. EXECUTIVE SUMMARY

This Clinical Outcome Assessment (COA) review is provided as a response to a request for consultation by the Division of Psychiatry Products (DPP) regarding NDA 207318. The sponsor used the Parkinson's Disease (PD) adapted Scale for the Assessment of Positive Symptoms (SAPS-PD), a clinician-reported outcome instrument administrated through patient semi structured interview, for the measurement of psychosis symptoms for use as a primary endpoint in a single phase 3 clinical trial in patients with adults 40 years or older with psychosis in Parkinson's disease. The sponsor's sought indication is for the treatment of psychosis associated with Parkinson's disease.

Psychosis in patients with Parkinson's disease is reported to have a clinical profile consisting of primarily of paranoid delusions and visual hallucinations that may be accompanied by other hallucinations. The SAPS-PD was adapted from a measure of psychosis in patients with schizophrenia to include the most common and relevant features of psychosis in Parkinson's disease. As a result, the SAPS-PD provides assessment of the two predominant symptoms in the target population (delusions and hallucinations). Patients with Parkinson's disease experience visual hallucinations more commonly than auditory hallucinations. However, while the SAPS-PD includes both types of hallucinations, it may give more weight to auditory hallucinations. In addition, other potential symptoms (e.g., illusions) (Ravina et al 2007, Fernandez et al 2008) do not appear in the SAPS-PD. While these limitations might affect the sensitivity to change of the SAPS-PD in the target patient population, we do not view them as critical flaws that would preclude the use of the SAPS-PD as a clinical outcome assessment to assess clinical benefit for regulatory use.

We also conclude that a 3-point change (out of 45) in the SAPS-PD does not clearly represent a clinically meaningful intra-patient change using anchor-based methods. Instead, we suggest that a minimal change of at least 5-7 points (out of 45) in this scale more clearly represents a clinically meaningful improvement.

While not a regulatory requirement, in the spirit of optimizing measurement for future clinical trials, we recommend further instrument development work be done including: investigation of whether the SAPS-PD is missing key psychosis symptoms such as illusions in Parkinson disease and confirmation of the adequacy of SAPS-PD using additional patient, caregiver or clinical expert input. The goal of this additional research is to confirm that the most important and relevant features are being assessed in a way that optimizes accuracy, reliability and ability to detect clinically meaningful change.

B. CLINICAL OUTCOME ASSESSMENT REVIEW

Materials reviewed:
Fernandez HH, Aarsland D, Fenelon G, et al. Scales to assess psychosis in Parkinson's disease: Critique and recommendations. Mov Disord. 2008;23(4):484-500.

Ravina B, Marder K, Fernandez HH, et al. Diagnostic Criteria for Psychosis in Parkinson's Disease: Report of an NINDS, NIMH Work Group. Mov Disord. 2007; 22(8):1061-1068.

Voss T, Bahr D, et al. Performance of a shortened Scale of Assessment of Positive Symptoms for Parkinson's disease psychosis. Parkinsonism Relat Disord. 2013;19(3): 295-299.

1 CONTEXT OF USE

1.1 Target Study Population and Clinical Setting

Adults aged 40 years or older with a clinical diagnosis of idiopathic Parkinson's disease for at least 1 year with psychotic symptoms that developed after the diagnosis of Parkinson's disease.

1.2 Clinical Trial Design, Protocol, and Analysis Plan

The applicant submitted a single phase 3, multi-center, placebo-controlled, double-blind trial to examine the safety and efficacy of pimavanserin in the treatment of psychosis in Parkinson's disease.

Inclusion criteria for the study:
Eligible subjects were males or females, aged 40 years or older, with a clinical diagnosis of idiopathic PD for at least 1 year with psychotic symptoms that developed after the diagnosis of PD and were present for at least 1 month before screening. The subject must have actively experienced psychotic symptoms each week during the month before screening. Psychotic symptoms included visual hallucinations and/or auditory hallucinations and/or delusions that were severe enough to warrant treatment with an antipsychotic agent. This was documented at screening by items A and B of the Neuropsychiatric Inventory (NPI), and defined as a score ≥4 on either the hallucinations (frequency x severity) or delusions (frequency x severity) scales or a total combined score (NPI-H+D) of ≥6. At baseline, subjects were required to have a SAPS-H or SAPS-D global item (H7 or D13) score ≥3 and a score ≥3 on at least one other non-global item using the SAPS-PD. At screening, subjects were required to have a Mini-Mental State Examination (MMSE) score ≥21 and be oriented to time and place. Subjects receiving anti-parkinson medications were required to have received stable doses for at least 1 month prior to baseline (Day 1) and during the study. Additionally, subjects were required to have a caregiver who provided consent, accompanied the subject to all study visits, and completed a questionnaire

to assess caregiver burden. Subjects and caregivers must have been willing and able to communicate in English for the purposes of the primary efficacy assessment, SAPS-PD.

The primary comparison for efficacy was the mean change in the SAPS-PD score from baseline (Day 1) to Day 43 between pimavanserin 40 mg and placebo analyzed using the mixed model repeated measures (MMRM) method for observed cases in the ITT analysis set. The ITT analysis set was the primary efficacy analysis set.

1.3 Endpoint Positioning

Primary Endpoint: The primary endpoint was the mean change in the SAPS-PD score from baseline (Day 1) to Day 43.

Analysis to support the primary endpoint included the percent change from baseline in the SAPS-PD, the SAPS-H+D scale (all 20 items), the percent change from baseline in the SAPS-H+D score, the domain scores for SAPS-H (7 items) and SAPS-D (13 items), the global rating item score for each domain (GSAPS-H and GSAPS-D), the sum of the 2 global scores (GSAPS-H+D), and the 20 individual SAPS-H+D item scores.

A remote rater (i.e., mental health evaluator) from the centralized service, MedAvante, conducted the SAPS rating in real-time using videoconference technology. The remote rater did not have access to the study design, entrance criteria, visit number, treatment assignment, or any study data for the subject or caregiver. A staff member and the subject's caregiver were present during the remote SAPS assessment.

Secondary Endpoints: The key secondary endpoint was the mean change from baseline in the combined Unified Parkinson's Disease Rating Scale (UPDRS) Parts II and III score (UPDRS Parts II+III) on Day 43, which was considered a measure of safety and function. UPDRS Part II and Part III component scores on Day 43 were also assessed. Other secondary endpoints included the Clinical Global Impression-Severity (CGI-S), CGI-Improvement (CGI-I), and CGI-I responders. The CGI was rated by a medically qualified clinician at the study center who did not have access to the SAPS data. Exploratory endpoints included the Scales for Outcomes in Parkinson's Disease (SCOPA)-sleep and the Caregiver Burden Scale (CBS).

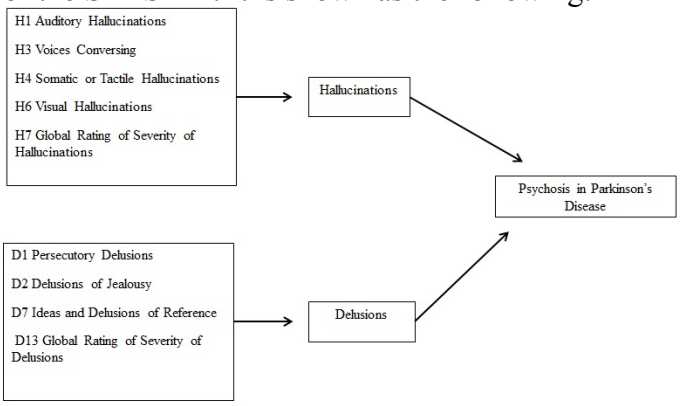
1.4 Proposed labeling or promotional claim(s) based on the COA

Treatment of psychosis associated with Parkinson's disease.

2 CONCEPT OF INTEREST AND CONCEPTUAL FRAMEWORK

The concept of interest is decrease in frequency of psychotic symptoms. The sponsor did not provide a conceptual framework of the SAPS-PD for review.

The reviewer constructed the putative conceptual framework of the SAPS-PD based on scoring of the SAPS-PD. It is shown as the following:



3 CLINICAL OUTCOME ASSESSMENT MEASURE(S)

The SAPS-PD is a 9-item instrument derived from the 20-item SAPS. The SAPS was developed as a clinician-reported outcome through a semi-structure interview with a patient. It was originally developed to study psychotic symptoms in schizophrenia patient population. The entire 20-item SAPS was administered to each patient and the 9 SAPS-PD items were extracted to form the SAPS-PD score. Caregivers were only interviewed if there were issues with during the interview with the patient or additional information was needed. The 9 items are:

- H1 Auditory Hallucinations
- H3 Voices Conversing
- H4 Somatic or Tactile Hallucinations
- H6 Visual Hallucinations
- H7 Global Rating of Severity of Hallucinations
- D1 Persecutory Delusions
- D2 Delusions of Jealousy
- D7 Ideas and Delusions of Reference

- D13 Global Rating of Severity of Delusions

The score of the SAPS-PD was a simple summation with a range of scores of 0-45 with a high score representing higher frequency of psychosis. The responses were on a 0 through 5 NRS, with 0=none; 1= unclear or questionable if the symptom is present , 2= symptom occurs 1 time in the past week (mild), 3= symptom occurs ***at least*** 2 times in the past week (moderate), 4= symptom occurs more days than not in the last week (marked), and 5= symptom occurs multiple times per day and is of notable duration (severe).

The SAPS semi-structure interview was conducted at baseline, Week 2, Week 4 and Week 6. The semi-structure interview reflect the past week of psychosis symptoms. An inter-rater correlation was established at 0.936 for study ACP-103-020.

4 CONTENT VALIDITY

The sponsor has not provided documentation of the content validity of the SAPS-PD for patients with psychosis for Parkinson's disease for review. Content validity is established from qualitative research and is defined as the extent to which the clinical outcome assessment instrument measures the concept of interest including evidence that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use. Qualitative research includes review of the current literature, concept elicitation and cognitive debriefing interviews with patients, caregivers and clinical experts. For the SAPS-PD, it is of interest to know if the SAPS-PD captures the relevant and important psychosis symptoms in Parkinson's disease and the recall period (i.e., past week) accurately captures the frequency of psychosis symptoms. In instrument development, it is important to establish content validity prior to evaluating the instrument's measurement properties and the instrument's ability to detect change. The *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*, is the optimal approach in selecting and/or developing a clinical outcome assessment that will best match a specific patient population.

The sponsor states that the content validity for the SAPS to be used in Parkinson's disease has been established from the support of the 2005 NINDS/NIHM consensus meeting and the 2005 Movement Disorder Society Task Force on Rating Scales in PD. The SAPS was developed for schizophrenia. Review of the Movement Disorder Society Task Force on Rating Scales in PD notes that the SAPS was not developed as an instrument to measure change; it is noted by the task force the SAPS does not rate the more common types of hallucinations or delusions in Parkinson's disease including illusions, and the hallucinations items are weighted toward auditory hallucinations. Visual hallucinations is more common in psychosis in Parkinson's disease.

The SAPS-PD was developed by modifying the SAPS based on principal component analysis and exploratory factor analysis using prior clinical trial data in psychosis in Parkinson's disease. SAPS items that were endorsed by <10% patients at baseline were excluded from the analyses to determine the modified SAPS-PD. It is unknown from the pooled clinical trial data how many people were included in the analyses. The <10% cut off was selected arbitrarily and was noted as a limitation by developers of the SAPS-PD. Input of clinical experts is described as above. No patients or caregivers provided input in the development of the SAPS-PD.

5 OTHER MEASUREMENT PROPERTIES (RELIABILITY, CONSTRUCT VALIDITY, ABILITY TO DETECT CHANGE)

The only information on the measurement properties of the SAPD-PD was inter-rater reliability. An inter-rater correlation was established at 0.936 for study ACP-103-020. Information on other measurement properties of the SAPS-PD was not provided for review.

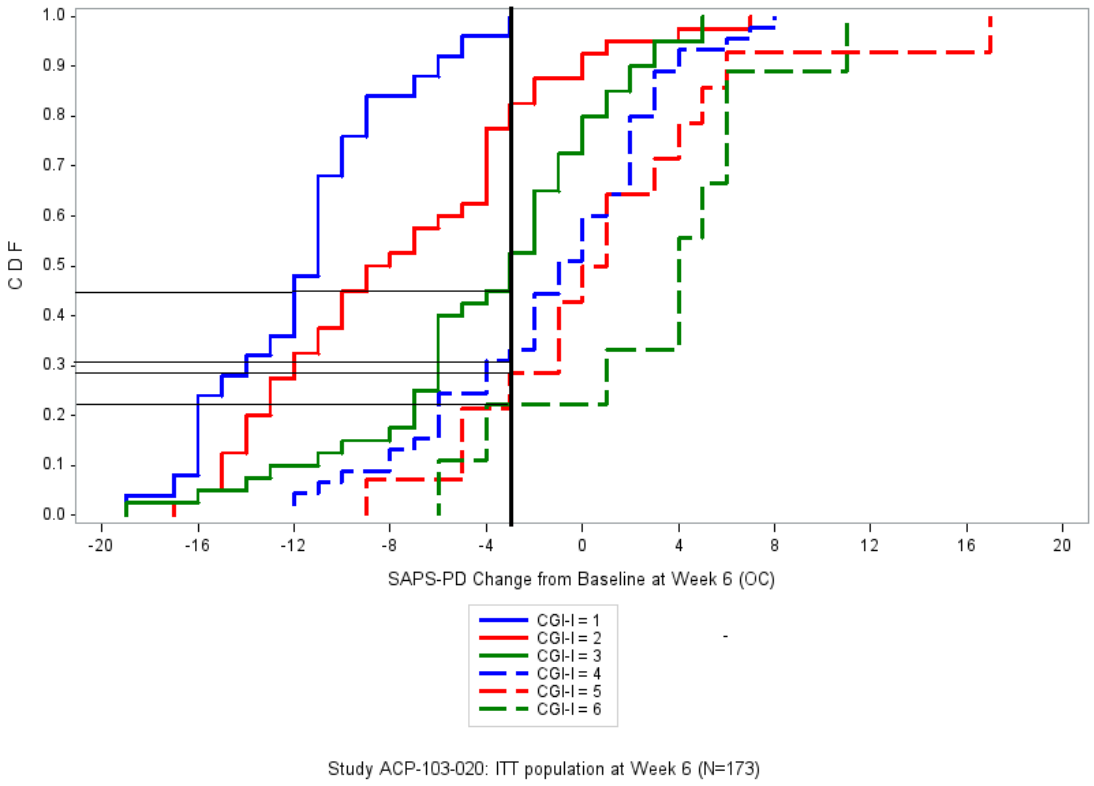
From review of the literature of the SAPS, the inter-rater reliability for SAPS summary score in psychotic patients is good (0.84). The intra-class coefficient (ICC) is 0.94. For the global domain, intra-class correlations ranged from 0.50 to 0.91. Test–retest reliability is weak–moderate (0.54). Internal consistency is weaker for the overall instrument (Cronbach α 0.48) than for the four global domain scores (ranging from 0.66 to 0.79). A single factor structure generally is not supported in the SAPS.

6 INTERPRETATION OF SCORES

Based on regression analysis described in the publication on the performance of the SAPS-PD (Voss et al., 2012), a clinically meaningful change defined as a 1-unit change in the Clinician Global Impression-Impact (CGI-I) scale is associated with a 2.33-point change in the SAPS-PD. A 1-unit change on CGI-I is consider a minimally improved intra-patient change on a 7-point CGI-I. the 7 units are: 1 = Very much improved, 2 = Much improved , 3 = Minimally improved, 4 = No Change, 5 = Minimally worse, 6 = Much worse, 7 = Very much worse
A 3-point change in the SAPS-PD for study ACP-103-020 represents the median of the SAPS-PD change score of the patients who showed minimal improvement (i.e., CGI-I=3) from baseline to Week 6 based on CGI-I assessed at Week 6. The median SAPS-PD change score of the patients rated as much improvement from baseline to Week 6 (CGI-I=2) is 7 points, as shown in the CDF curves below (Figure 1). The CDF curves also show that there is little separation between minimal improvement (CGI-I =3), no change (CGI-I=4), minimally worse (CGI-I=5), and much worse (CGI-I=6). They show that large percentages of no change and worsen patients also had ≥ 3-point change in SAPS-PD (i.e., 44%, 31%, 29%, and 22% for minimally improved, no change, minimally worse, and much worse, respectively). That is, there is a certain amount of

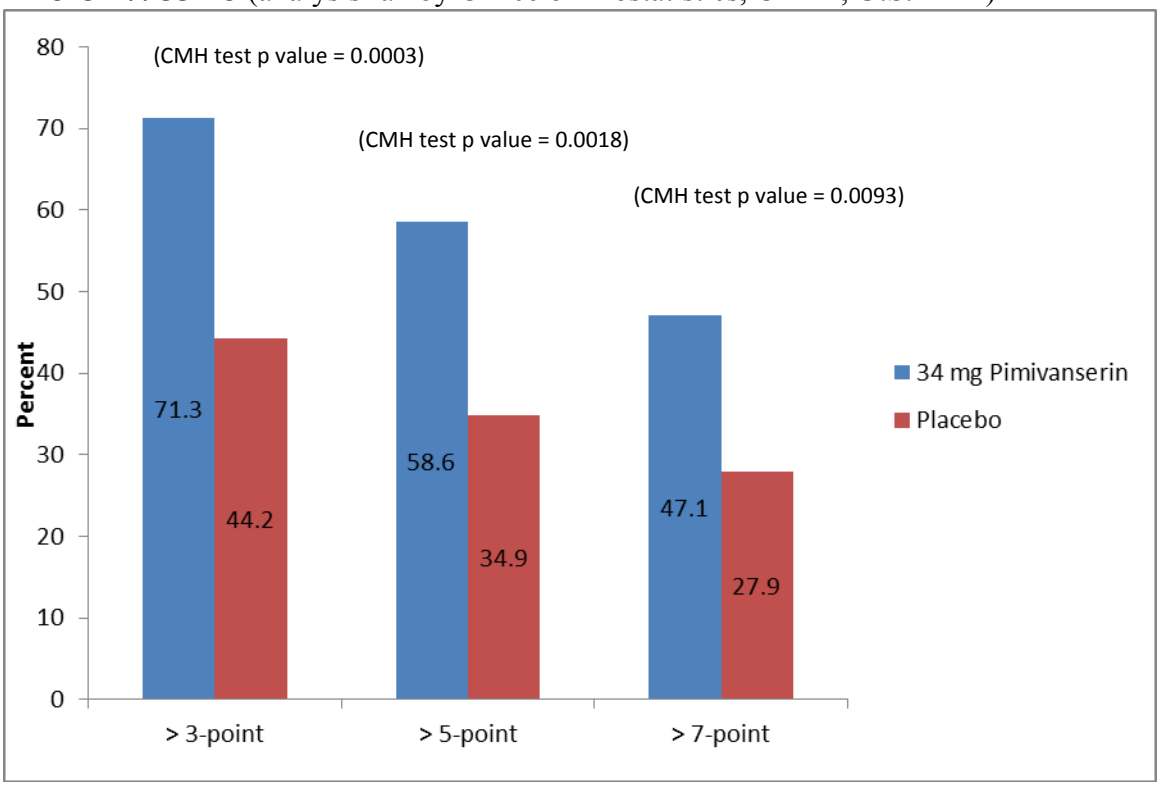
noise (uncertainty) of using the 3-point change as the threshold. In this regard, a larger threshold that represents clinically meaningful improvement with higher certainty, such as 7-point or 5-point change, may be considered.

Figure 1. SAPS-PD by CGI level (analysis ran by Office of Biostatistics, CDER, U.S. FDA)



The histogram (Figure 2) below show the percentages of patients meeting the 3-, 5-, and 7-points changes of the two treatment arms. The histogram shows that the 3-point threshold is a low estimate of clinically meaningful change as 44.2 % of the patients in the placebo arm had more the 3-point change in the SAPS-PD total score.

Figure 2. Proportion of Patients who had SAPS-PD score improvement at the End of Week 6 (analysis ran by Office of Biostatistics, CDER, U.S. FDA)



A 5 to 7-point change may represent a reasonable meaningful improvement.

7 LANGUAGE TRANSLATION AND CULTURAL ADAPTATION

N/A. The trial was conducted at 66 centers (63 in the US and 3 in Canada) and the ability to communicate in English was an inclusion criteria.

8 REFORMATTING FOR NEW METHOD OR MODE OF ADMINISTRATION

The complete SAPS was administered through a centralized rater service as a semi-structure interview. This includes the 9-item SAPS-PD.

.

9 REVIEW USER MANUAL

A central rating provider was used for the administration of the SAPS of all subjects. Standard procedures were established at the beginning of the study for training and calibration of individuals responsible for training and monitoring each rater. Training and calibration of raters as well as monitoring of centrally-based clinical raters was provided in the training manual for raters. All trainers and raters held at least a Master's level degree with training in Psychology, Social Work, or Medicine, and were experienced in administration of the SAPS or similar scales (e.g., Positive and Negative Syndrome Scale [PANSS]). The training and user manual was consistent with the concept of interest of the instrument and its context of use. An inter-rater correlation was established at 0.936 for study ACP-103-020.

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

MICHELLE L CAMPBELL
02/22/2016

WEN-HUNG CHEN
02/22/2016

ELEKTRA J PAPADOPOULOS
02/23/2016

Summary Pathology Comments:

The 9 laboratory animal studies (5 rat, 2 mouse, 2 monkey) listed above were conducted over a period of 12 years (2002-2014) at 4 different laboratories/locations (Quintiles UK, SNBL, Quintiles Kansas City, WIL) and read by different pathologists. Most study narratives read as though the study pathologists / authors were not familiar with study findings recorded in studies conducted earlier. This was evidenced by variation in terminology and interpretation of treatment related changes.
Our evaluation revealed diagnoses compatible with or implying **multi-systemic PLD** in 7/9 studies (R1/1, MK 3/1, R3/1, MS14D/0, MK12/4, R24/0, R6/6). Studies negative for PLD were the MS 3/1 study (which resulted in non-PLD induced liver toxicity at ≥10 mg) and the R6/3 study for which a treatment related increase in alveolar macrophages was reported in females at the 30 mg/kg level. TEM is in our opinion necessary to unequivocally rule out a low grade PLD manifestation at this level for this study. TEM confirming PLD was conducted only for tissues from one (R6/6) of the 7 PLD positive studies. In one of the 7 studies positive for PLD (MK12/4), the term PLD was never specifically stated. To us, however, the description of foamy macrophage / cytoplasm of this study was compatible with multi-systemic PLD.

Female rats appeared to be more sensitive than males when comparing overall incidence tables for PLD related changes. 2 of the 7 PLD positive studies specifically stated that rat lung and kidneys were generally more severely affected, when multiple organs were reported with PLD. In all PLD positive studies, macrophages were characterized by some form of dose dependent increases in cytoplasmic vacuolation (**foamy macrophages**). In 4 (R1/1, R3/1, R24/0, R6/6) of 7 PLD positive studies, **eosinophilic material** (generally interpreted to be phospholipid) was observed inside vacuolated macrophage as well as extracellularly within the alveolar lumen (interpreted to be a consequence of macrophage lysis). Only in the R6/6 study, macrophages were further described to also include multinucleated giant cells and cholesterol clefts. Multinucleated giant cells generally result from macrophage fusion secondary to an inability of the macrophage to digest phagocytosed material; intracellular cholesterol clefts in macrophages are indicative of lipid rich materials stored within the macrophages. The manifestation of both of these features is theoretically conceivable considering the drug-class context of phospholipidosis. Multinucleated giant cells as well as cholesterol clefts are, however, not a typical feature of phospholipidosis; therefore these observations may warrant further investigation and safety consideration.

In the rat study of the longest duration (R24/0) **inflammatory cell infiltrates and inflammation** were described to be **associated** with the vacuolated macrophages and the extracellular material. In addition the R1/1 and the R3/1 also recorded inflammatory cell responses although the combined inflammatory response of the R1/1 study did not necessarily show incidences indicating a treatment relationship. Nevertheless, overall it appears that with PLD at higher doses, alveolar macrophages and extracellular material elicit a low grade inflammatory response in the pulmonary parenchyma. The incidences of inflammatory responses generally correlated to the higher PLD severity scores and treatment doses of Pimavanserin. In one of these rat studies (R3/1 – following the correction of an assumed summary table compilation error (see pathology comment for study R3/1 for details), the recorded inflammatory response was restricted to the 1 month recovery group and consisted of “adenomatous [type 2 pneumocyte] hyperplasia in males and **chronic inflammation** in females”. Chronic inflammation by definition implies low degrees of interstitial **collagen deposition which is consistent with fibrosis**, resulting from long standing (weeks to months) inflammatory processes. In the recovery group of the R6/6 study, minimal to mild **interstitial and pleural fibrosis** was diagnosed which was identified by

the study authors as a permanent change. Resolution of chronic inflammation with low grade collagen deposition, as a component of chronic inflammation, will result in small areas of focal to multifocal fibrosis which persists while the inflammatory cellular components subside over time.

Fibrosis:
Optimally a slide review would have been conducted by CFSAN Pathology to assess the quality and quantity of the specific histological changes associated with the reported PLD. However, in lieu of slides, we concluded based on the overall information provided in the studies above, that the described '**fibrosis**' appears different from primary pulmonary fibrosis and is not compatible with "human pulmonary fibrosis".
The described changes are not suggestive of the spectrum of pathologic changes usually associated with the group of chronic diffuse lung disorders or acute lung injury associated with adverse drug reactions in humans. We propose a PLD process with an associated low grade ongoing inflammatory cell response which organizes over time (chronicity) resulting in collagen deposits manifesting as fibrosis". This "fibrosis" is a minor component of the lesions and is interpreted as being a secondary consequence of the inflammatory reaction. Fibrosis (newly produced collagen) at very small amounts is difficult to discern histologically in an H&E stained slide from preexisting collagen as both stain eosinophilic (pink). To more readily identify and visualize the degree of fibrosis, a special stain (Masson's trichrome) for collagen is generally used.

Your specific Questions: reference 1 a

1. *The sponsor acknowledges that pimavanserin causes widespread "systemic" phospholipidosis in mice, rats and monkeys. They also stated in their toxicology summary section that fibrosis occurred in the lungs of rats and that the finding was considered a permanent change that is toxicologically relevant to humans. However, the sponsor suggests that the finding is rat-specific, only dose- and not duration-dependent and there is an adequate safety margin compared to human exposures (9-fold). Do you agree with the sponsor's conclusions that the lung fibrosis is not a relevant risk to humans?*

Pathology comment: With regard to PLD and fibrosis observed in the lungs of rats the sponsor states the following (reference 1 f comment #2):

> "In the rat study (Study WIL-616007), putative events leading to the observed lung fibrosis, i.e., collections of large foamy macrophages and the presence of extracellular material with or without chronic inflammation and eventual fibrosis [ ], reflect a known path in fibrogenesis.

We agree with the sponsor's assessment regarding the underlying pathomechanism leading to the fibrosis described in these studies (secondary to chronic inflammation which in turn is a response to PLD).

The sponsor proceeds (comment #2):

> "It is important to note that the moderate to severe phospholipidosis, the putative initiating factor for the lesion, is dose related but not duration related."

We agree with the sponsor that the PLD appears to be dose dependent, evidenced by e.g. the R6/6 study showing reduced incidences and severities of the PLD in the 60 mg/kg dose compared to the 90 mg/kg group of both sexes. However, we disagree with the sponsor that the PLD is not duration related. While PLD changes are not reported for males at the 30 mg/kg dose in the R3/1 and R6/3 studies, this dose level is affected by PLD after prolonged treatment with Pimavanserin in males of the R24/0 study. In addition in our opinion, multi-systemic PLD is not rat specific as it occurs in multiple species (mouse, monkey and rat). The manifestation of the type of fibrosis observed (secondary to inflammation) is not rat specific either but depends on the severity of the PLD and the degree and chronicity of the inflammation the PLD is associated with.

With regard to relevant risk, the sponsor stated (comment #2):

> "The observation of fibrosis is toxicologically relevant to humans; however, the finding in rats occurred only at a high dose (a dose causing lethality), and at ~18-fold human exposure"…[…….] Fibrosis in rats occurred only in animals that had moderate to severe phospholipidosis that was slow to resolve and only at high doses. With a high margin (>15 fold) for the fibrotic finding in rat, this is not considered to be a concern for patients".

We do not see that the sponsor specifically states that "*the lung fibrosis is not a relevant risk to humans?"* We do agree with the sponsor, that the observed minimal multifocal fibrosis that resides following longstanding low grade inflammation in response to PLD at high doses is

relevant to humans. The exposure margins and resulting concern for patients, are depending on the assessment of the adverse effect levels and are beyond the scope of this evaluation. Events considered adverse secondary to PLD reported in some of the 9 studies evaluated are inflammation (including chronic inflammation with fibrosis) and type 2-pneumocyte hyperplasia (Nikula et al., 2013; STP Position Paper: Interpreting the significance of increased alveolar macrophages in rodents following inhalation of pharmaceutical materials).

2. *In your expert opinion, do you have a hypothesis for why the lung fibrosis was primarily observed at the end of the 6-month recovery period and not in main study animals?*

Pathology comment: as outlined above under the Pathology summary comments we agree with the sponsors scenario outlined in reference 1 f comment #3, that a continuum underlies the development of the minimal-mild, multifocal fibrosis observed in the R6/6 study consisting of PLD (at high doses with multinucleated giant cells and cholesterol clefts), extracellular material, inflammation, and chronic inflammation (with fibrosis). The fibrosis reported in the R6/6 study represents a "point in time" observation which after cessation of treatment at which the resolution process resulted in remnant minimal-mild, multifocal foci of fibrosis while the originally associated inflammation had resolved.

3. *In your expert opinion, is it accurate to conclude that lung fibrosis is progression of persistent inflammation which in turn is a result of PLD? If so, is it reasonable to conclude that lung fibrosis is a clinically relevant outcome if pimavanserin is to be administered over a long period of time?*

Pathology comment: The fibrosis findings presented in the studies evaluated are in our opinion not consistent with the term "lung fibrosis" which implies a primary fibrotic process. We prefer the term "chronic inflammation" which implies some degree of organizing fibrosis or "fibrosis secondary to chronic inflammation". The level of secondary inflammatory fibrosis reported in the rat were described as of minimal to mild severity and being focal to multifocal in distribution; therefore, the clinical relevance of this finding depends on dose and duration. The relevant clinical parameter to monitor for is inflammation. We do consider any PLD **with** inflammation (including chronic inflammation) a clinically relevant outcome.

4. *Although the sponsor did not conduct any specific pulmonary function tests in animals, do the clinical respiratory-related findings in rats, along with the macro- and microscopic findings in the lungs suggest that pimavanserin may adversely affect lung function? If so, would this be a clinically relevant finding that can be monitorable?*

Pathology comment: The studies evaluated reported in 3/5 rat studies respiratory related findings of "noisy breathing" (R1/1) or "rales" (R24/0, R6/6) which correlated to pulmonary PLD **and** the amount of extracellular eosinophilic material. Auscultation of rales would be a clinically monitorable parameter. Monkeys and mice were not reported to show rales or noisy

breathing. The clinical relevance of this observation would, however, better be answered by a clinician.

5. Please provide any additional comments/recommendations you may have regarding the significance of the histopathological findings of pimavanserin in animals.

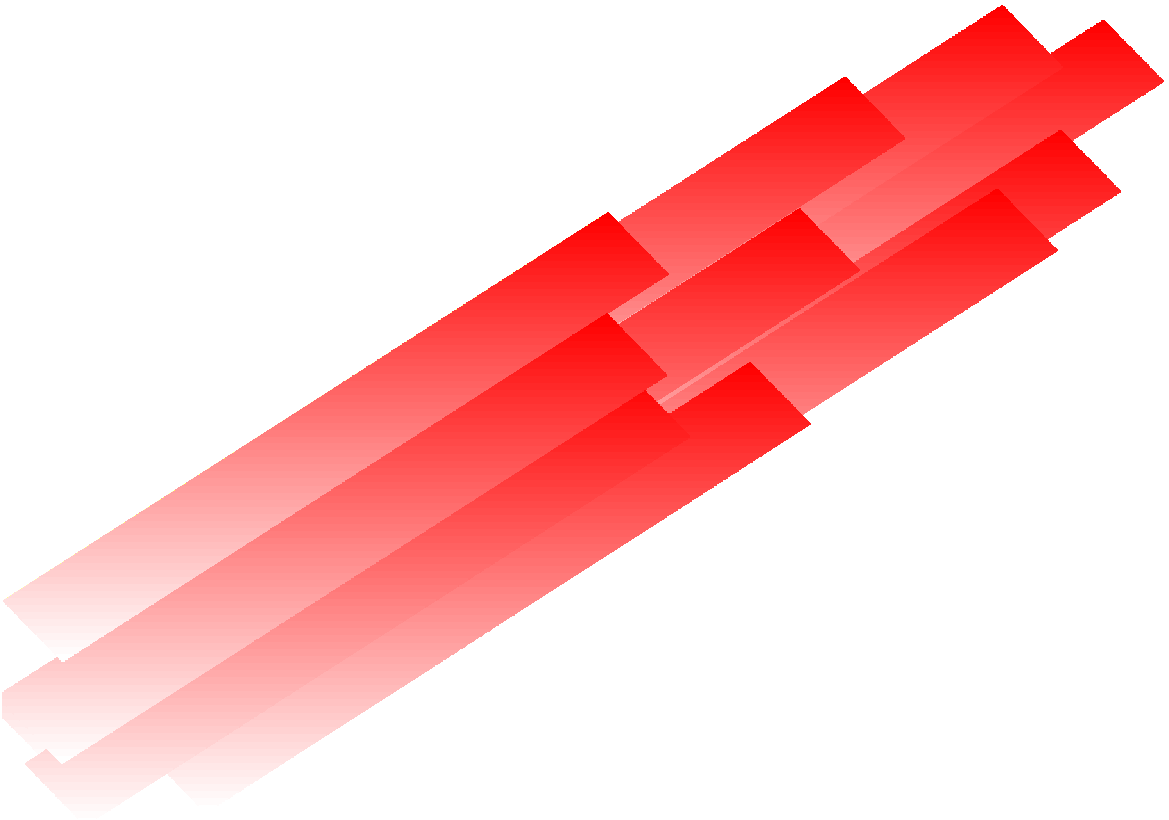
Pathology comment:
The lung and kidney appeared to be the most sensitive organs in the rat. In 2/5 rat studies (R3/1, R6/6), renal changes were reported consistent with “renal nephrosis” (tubular degeneration) in response to PLD with renal tubular vacuolation. For the R6/6 study we concurred with the original study pathologist, that the treatment related renal findings described are consistent with a renal manifestation, but not a pre-renal manifestation, as amended in the overall study report.

Please let us know if you have any questions.

Sabine Francke, D.V.M., Ph.D., FIATP and Steven Mog D.V.M., DACVP

Guidance for Industry

Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
May 1998
Clinical 6

Guidance for Industry

Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products

Additional copies are available from:
the Drug Information Branch (HFD-210),
Center for Drug Evaluation and Research (CDER),
5600 Fishers Lane, Rockville, MD 20857 (Tel) 301-827-4573
Internet at http://www.fda.gov/cder/guidance/index.htm
or
Office of Communication,
Training, and Manufacturers Assistance (HFM-40)
Center for Biologics Evaluation and Research (CBER)
1401 Rockville Pike, Rockville, MD 20852-1448
http://www.fda.gov/cber/guidelines.htm
(Fax) 888-CBERFAX or 301-827-3844
(Voice Information) 800-835-4709 or 301-827-1800

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
May 1998
Clinical 6

TABLE OF CONTENTS

GUIDANCE FOR INDUSTRY[1]

Providing Clinical Evidence of Effectiveness[2] for Human Drug and Biological Products

I. INTRODUCTION

This document is intended to provide guidance to applicants planning to file new drug applications (NDAs), biologics license applications (BLAs), or applications for supplemental indications on the evidence to be provided to demonstrate effectiveness.

This document is also intended to meet the requirements of subsections 403(b)(1) and (2) of the Food and Drug Administration Modernization Act (the Modernization Act) of 1997 for human drug and biological products (P.L. 105-115).[3] Subsection 403(b)(1) directs FDA to provide guidance on the circumstances in which published matter may be the basis for approval of a supplemental application for a new indication. Section III of this guidance satisfies this requirement by describing circumstances in which published matter may partially or entirely support approval of a supplemental application. Subsection 403(b)(2) directs FDA to provide guidance on data requirements that will avoid duplication of previously submitted data by recognizing the availability of data previously submitted in support of an original application to support approval of a supplemental application. Section II of this guidance satisfies this requirement by describing a range of circumstances in which related existing data, whether from an original application or other sources, may be used to support approval of a supplemental application.

In 1962, Congress amended the Federal Food, Drug, and Cosmetic Act to add a requirement that, to obtain marketing approval, manufacturers demonstrate the effectiveness of their products through the conduct of adequate and well-controlled studies. Since then, the issue of what constitutes sufficient evidence of effectiveness has been debated by the Agency, the scientific community, industry, and others. Sound evidence of effectiveness is a crucial component of the Agency's benefit-risk assessment of a new product or use. At the same time, the demonstration of effectiveness represents a major component of drug development time and cost; the amount

---

[1] This guidance document represents the agency's current thinking on providing clinical evidence of effectiveness for human drug and biological products. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both.

[2] As used in this guidance, the term efficacy refers to the findings in an adequate and well-controlled clinical trial or the intent of conducting such a trial and the term effectiveness refers to the regulatory determination that is made on the basis of clinical efficacy and other data.

[3] The Modernization Act requirements in Section 403 also apply to animal drugs and medical devices. These products will be addressed in separate guidances.

and nature of the evidence needed can therefore be an important determinant of when and whether new therapies become available to the public. The public health is best served by the development of sound evidence of effectiveness in an efficient manner.

The science and practice of drug development and clinical evaluation have evolved significantly since the effectiveness requirement for drugs was established, and this evolution has implications for the amount and type of data needed to support effectiveness in certain cases. As a result of medical advances in the understanding of pathogenesis and disease staging, it is increasingly likely that clinical studies of drugs will be more narrowly defined to focus, for example, on a more specific disease stage or clinically distinct subpopulation. As a consequence, product indications are often narrower, the universe of possible indications is larger, and data may be available from a number of studies of a drug in closely related indications that bear on a determination of its effectiveness for a new use. Similarly, there may be studies of a drug in different populations, studies of a drug alone or in combination, and studies of different doses and dosage forms, all of which may support a particular new use of a drug. At the same time, progress in clinical evaluation and clinical pharmacology have resulted in more rigorously designed and conducted clinical efficacy trials, which are ordinarily conducted at more than one clinical site. This added rigor and scope has implications for a study's reliability, generalizability, and capacity to substantiate effectiveness.

Given this evolution, the Agency has determined that it would be appropriate to articulate its current thinking concerning the quantitative and qualitative standards for demonstrating effectiveness of drugs and biologics. FDA hopes that this guidance will enable sponsors to plan drug development programs that are sufficient to establish effectiveness without being excessive in scope. The guidance should also bring greater consistency and predictability to FDA's assessment of the clinical trial data needed to support drug effectiveness.

Another major goal of this guidance is to encourage the submission of supplemental applications to add new uses to the labeling of approved drugs. By articulating how it currently views the quantity and quality of evidence necessary to support approval of a new use of a drug, FDA hopes to illustrate that the submission of supplements for new uses need not be unduly burdensome.

II. QUANTITY OF EVIDENCE NECESSARY TO SUPPORT EFFECTIVENESS

A. Legal Standards for Drug and Biological Products

Drugs: The effectiveness requirement for drug approval was added to the Federal Food, Drug, and Cosmetic Act (the Act or the FDC Act) in 1962. Between passage of the Act in 1938 and the 1962 amendments, drug manufacturers were required to show only that their drugs were safe. The original impetus for the effectiveness requirement was Congress's growing concern about the misleading and unsupported claims being made by pharmaceutical companies about their drug products coupled with high drug prices. After two years of hearings on these issues, Congress adopted the 1962 Drug Amendments,

which included a provision requiring manufacturers of drug products to establish a drug's effectiveness by "substantial evidence." *Substantial evidence* was defined in section 505(d) of the Act as "evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof."

Since the 1962 Amendments added this provision to the statute, discussions have ensued regarding the quantity and quality of the evidence needed to establish effectiveness. With regard to quantity, it has been FDA's position that Congress generally intended to require at least two adequate and well-controlled studies, each convincing on its own, to establish effectiveness. (See e.g., Final Decision on Benylin, 44 FR 51512, 518 (August 31, 1979); *Warner-Lambert Co. V. Heckler*, 787 F. 2d 147 (3d Cir. 1986)). FDA's position is based on the language in the statute[4] and the legislative history of the 1962 amendments. Language in a Senate report suggested that the phrase "adequate and well-controlled investigations" was designed not only to describe the quality of the required data but the "quantum" of required evidence. (S. Rep. No. 1744, Part 2, 87th Cong. 2d Sess. 6 (1962))

Nevertheless, FDA has been flexible within the limits imposed by the congressional scheme, broadly interpreting the statutory requirements to the extent possible where the data on a particular drug were convincing. In some cases, FDA has relied on pertinent information from other adequate and well-controlled studies of a drug, such as studies of other doses and regimens, of other dosage forms, in other stages of disease, in other populations, and of different endpoints, to support a single adequate and well-controlled study demonstrating effectiveness of a new use. In these cases, although there is only one study of the exact new use, there are, in fact, multiple studies supporting the new use, and expert judgment could conclude that the studies together represent substantial evidence of effectiveness. In other cases, FDA has relied on only a single adequate and well-controlled efficacy study to support approval — generally only in cases in which a single multicenter study of excellent design provided highly reliable and statistically strong evidence of an important clinical benefit, such as an effect on survival, and a confirmatory study would have been difficult to conduct on ethical grounds.

In section 115(a) of the Modernization Act, Congress amended section 505(d) of the Act to make it clear that the Agency may consider "data from one adequate and well-controlled clinical investigation and confirmatory evidence" to constitute substantial

[4] Section 505(d) of the Act uses the plural form in defining "substantial evidence" as "adequate and well-controlled investigations, including clinical investigations." See also use of "investigations" in section 505(b) of the Act, which lists the contents of a new drug application.

evidence if FDA determines that such data and evidence are sufficient to establish effectiveness. In making this clarification, Congress confirmed FDA's interpretation of the statutory requirements for approval and acknowledged the Agency's position that there has been substantial progress in the science of drug development resulting in higher quality clinical trial data.

Biologics. Biological products are approved under authority of section 351 of the Public Health Service Act (PHS Act) (42 U.S.C.§ 262). Under section 351, as in effect since 1944, licenses for biologics have been issued only upon a showing that the products meet standards designed to ensure the "continued safety, purity, and potency" of the products. *Potency* has long been interpreted to include effectiveness (21 CFR 600.3(s)). In 1972, FDA initiated a review of the safety and effectiveness of all previously licensed biologics. The Agency stated then that proof of effectiveness would consist of controlled clinical investigations as defined in the provision for "adequate and well-controlled studies" for new drugs (21 CFR 314.126), unless waived as not applicable to the biological product or essential to the validity of the study when an alternative method is adequate to substantiate effectiveness (21 CFR 601.25 (d) (2)). One such adequate alternative was identified to be serological response data where a previously accepted correlation with clinical effectiveness exists. As with nonbiological drug products, FDA has approved biological products based on single, multicenter studies with strong results.

Although section 123(a) of the Modernization Act amended section 351 of the PHS Act to make it clear that separate licenses are not required for biological products and the establishments at which the products are made, the evidentiary standard for a biological product was not changed: the product must be shown to be "safe, pure, and potent" (section 351 (a)(2) of the PHS Act as amended). In the Modernization Act (section 123(f)) Congress also directed the agency to take measures to "minimize differences in the review and approval" of products required to have approved BLAs under section 351 of the PHS Act and products required to have approved NDAs under section 505(b)(1) of the FDC Act.

B. Scientific Basis for the Legal Standard

The usual requirement for more than one adequate and well-controlled investigation reflects the need for *independent substantiation* of experimental results. A single clinical experimental finding of efficacy, unsupported by other independent evidence, has not usually been considered adequate scientific support for a conclusion of effectiveness. The reasons for this include the following.

- Any clinical trial may be subject to unanticipated, undetected, systematic biases. These biases may operate despite the best intentions of sponsors and investigators, and may lead to flawed conclusions. In addition, some investigators may bring conscious biases to evaluations.

- The inherent variability in biological systems may produce a positive trial result by chance alone. This possibility is acknowledged, and quantified to some extent, in the statistical evaluation of the result of a single efficacy trial. It should be noted, however, that hundreds of randomized clinical efficacy trials are conducted each year with the intent of submitting favorable results to FDA. Even if all drugs tested in such trials were ineffective, one would expect one in forty of those trials to "demonstrate" efficacy by chance alone at conventional levels of statistical significance.[5] It is probable, therefore, that false positive findings (i.e., the chance appearance of efficacy with an ineffective drug) will occur and be submitted to FDA as evidence of effectiveness. Independent substantiation of a favorable result protects against the possibility that a chance occurrence in a single study will lead to an erroneous conclusion that a treatment is effective.

- Results obtained in a single center may be dependent on site or investigator specific factors (e.g., disease definition, concomitant treatment, diet). In such cases, the results, although correct, may not be generalizable to the intended population. This possibility is the primary basis for emphasizing the need for independence in substantiating studies.

- Rarely, favorable efficacy results are the product of scientific fraud.

Although there are statistical, methodologic, and other safeguards to address the identified problems, they are often inadequate to address these problems in a single trial. Independent substantiation of experimental results addresses such problems by providing consistency across more than one study, thus greatly reducing the possibility that a biased, chance, site-specific, or fraudulent result will lead to an erroneous conclusion that a drug is effective.

The need for independent substantiation has often been referred to as the need for replication of the finding. Replication may not be the best term, however, as it may imply that precise repetition of the same experiment in other patients by other investigators is the only means to substantiate a conclusion. Precise replication of a trial is only one of a number of possible means of obtaining independent substantiation of a clinical finding and, at times, can be less than optimal as it could leave the conclusions vulnerable to any systematic biases inherent to the particular study design. Results that are obtained from studies that are of different design and independent in execution, perhaps evaluating different populations, endpoints, or dosage forms, may provide support for a conclusion of effectiveness that is as convincing as, or more convincing than, a repetition of the same study.

[5] p-value = 0.05, two-tailed, which implies an error rate in the efficacy (false positive) tail of 0.025 or one in forty.

C. The Quantity of Evidence to Support Effectiveness

The following three sections provide guidance on the quantity of evidence needed in particular circumstances to establish substantial evidence of effectiveness. Section 1 addresses situations in which effectiveness of a new use may be extrapolated entirely from existing efficacy studies. Section 2 addresses situations in which a single adequate and well-controlled study of a specific new use can be supported by information from other related adequate and well-controlled studies, such as studies in other phases of a disease, in closely related diseases, of other conditions of use (different dose, duration of use, regimen), of different dosage forms, or of different endpoints. Section 3 addresses situations in which a single multicenter study, without supporting information from other adequate and well-controlled studies, may provide evidence that a use is effective.

In each of these situations, it is assumed that any studies relied on to support effectiveness meet the requirements for adequate and well-controlled studies in 21 CFR 314.126. It should also be appreciated that reliance on a single study of a given use, whether alone or with substantiation from related trial data, leaves little room for study imperfections or contradictory (nonsupportive) information. In all cases, it is presumed that the single study has been appropriately designed, that the possibility of bias due to baseline imbalance, unblinding, post-hoc changes in analysis, or other factors is judged to be minimal, and that the results reflect a clear prior hypothesis documented in the protocol. Moreover, a single favorable study among several similar attempts that failed to support a finding of effectiveness would not constitute persuasive support for a product use unless there were a strong argument for discounting the outcomes in the studies that failed to show effectiveness (e.g., study obviously inadequately powered or lack of assay sensitivity as demonstrated in a three-arm study by failure of the study to show efficacy of a known active agent).

Whether to rely on a single study to support an effectiveness determination is not often an issue in contemporary drug development. In most drug development situations, the need to find an appropriate dose, to study patients of greater and lesser complexity or severity of disease, to compare the drug to other therapy, to study an adequate number of patients for safety purposes, and to otherwise know what needs to be known about a drug before it is marketed will result in more than one adequate and well-controlled study upon which to base an effectiveness determination.

This guidance is not intended to provide a complete listing of the circumstances in which existing efficacy data may provide independent substantiation of related claims; rather, it provides examples of the reasoning that may be employed. The examples are applicable whether the claim arises in the original filing of an NDA or BLA, or in a supplemental application.

1. Extrapolation from Existing Studies

In certain cases, effectiveness of an approved drug product for a new indication, or effectiveness of a new product, may be adequately demonstrated without additional adequate and well-controlled clinical efficacy trials. Ordinarily, this will be because other types of data provide a way to apply the known effectiveness to a new population or a different dose, regimen or dosage form. The following are examples of situations in which effectiveness might be extrapolated from efficacy data for another claim or product.

a. Pediatric uses

The rule revising the Pediatric Use section of product labeling (21 CFR 201.57(f)(9)(iv)) makes allowance for inclusion of pediatric use information in labeling without controlled clinical trials of the use in children. In such cases, a sponsor must provide other information to support pediatric use, and the Agency must conclude that the course of the disease and the effects of the drug are sufficiently similar in the pediatric and adult populations to permit extrapolation from adult efficacy data to pediatric patients. Evidence that could support a conclusion of similar disease course and similar drug effect in adult and pediatric populations includes evidence of common pathophysiology and natural history of the disease in the adult and pediatric populations, evidence of common drug metabolism and similar concentration-response relationships in each population, and experience with the drug, or other drugs in its therapeutic class, in the disease or condition or related diseases or conditions. Examples in which pediatric use labeling information has been extrapolated from adult efficacy data include ibuprofen for pain and loratidine for seasonal allergic rhinitis.

b. Bioequivalence

The effectiveness of alternative formulations and new dosage strengths may be assessed on the basis of evidence of bioequivalence.

c. Modified-release dosage forms

In some cases, modified release dosage forms may be approved on the basis of pharmacokinetic data linking the new dosage form to a previously studied immediate-release dosage form. Because the pharmacokinetic patterns of modified-release and immediate-release dosage forms are not identical, it is generally important to have some understanding of the relationship of blood concentration to response, including an understanding of the time course of that relationship, to extrapolate the immediate-release

data to the modified-release dosage form.

d. Different doses, regimens, or dosage forms

Dose-response relationships are generally continuous such that information about the effectiveness of one dose, dosage regimen, or dosage form is relevant to the effectiveness of other doses, regimens, or dosage forms. Where blood levels and exposure are not very different, it may be possible to conclude that a new dose, regimen, or dosage form is effective on the basis of pharmacokinetic data alone. Even if blood levels are quite different, if there is a well-understood relationship between blood concentration and response, including an understanding of the time course of that relationship, it may be possible to conclude that a new dose, regimen, or dosage form is effective on the basis of pharmacokinetic data without an additional clinical efficacy trial. In this situation, pharmacokinetic data, together with the well-defined pharmacokinetic/pharmacodynamic (PK/PD) relationship, are used to translate the controlled trial results from one dose, regimen, or dosage form to a new dose, regimen, or dosage form (See also section II.C.2.a).

2. Demonstration of Effectiveness by a Single Study of a New Use, with Independent Substantiation From Related Study Data

The discussion that follows describes specific examples in which a single study of a new use, with independent substantiation from study data in related uses, could provide evidence of effectiveness. In these cases, the study in the new use and the related studies support the conclusion that the drug has the effect it is purported to have. Whether related studies are capable of substantiating a single study of a new use is a matter of judgment and depends on the quality and outcomes of the studies and the degree of relatedness to the new use.

a. Different doses, regimens, or dosage forms

As discussed in Sections II.C.1.d, it may be possible to conclude that a new dose, regimen, or dosage form is effective on the basis of pharmacokinetic data without an additional clinical efficacy trial where blood levels and exposure are not very different or, even if quite different, there is a well-understood relationship between blood concentration and response. Where the relationship between blood concentration and response is not so well understood and the pharmacokinetics of the new dose, regimen, or dosage form differ from the previous one, clinical efficacy data will likely be necessary to support effectiveness of a new regimen. In this case, a single additional efficacy study should ordinarily be sufficient. For example, a single controlled trial was needed to support the recent approval of a once

daily dose of risperidone because the once daily and twice daily regimens had different pharmacokinetics and risperidone's PK/PD relationship was not well understood.

b. Studies in other phases of the disease

In many cases, therapies that are effective in one phase of a disease are effective in other disease phases, although the magnitude of the benefit and benefit-to-risk relationship may differ in these other phases. For example, if a drug is known to be effective in patients with a refractory stage of a particular cancer, a single adequate and well-controlled study of the drug in an earlier stage of the same tumor will generally be sufficient evidence of effectiveness to support the new use.

c. Studies in other populations

Often, responses in subsets of a particular patient population are qualitatively similar to those in the whole population. In most cases, separate studies of effectiveness in demographic subsets are not needed (see also discussion of the pediatric population in section II.C.1.a) However, where further studies are needed, a single study would ordinarily suffice to support effectiveness in age, race, gender, concomitant disease, or other subsets for a drug already shown to be generally effective in a condition or to be effective in one population. For example, a single study was sufficient to support tamoxifen use in breast cancer in males.

d. Studies in combination or as monotherapy

For a drug known to be effective as monotherapy, a single adequate and well-controlled study is usually sufficient to support effectiveness of the drug when combined with other therapy (as part of a multidrug regimen or in a fixed-dose combination). Similarly, known effectiveness of a drug as part of a combination (i.e., its contribution to the effect of the combination is known) would usually permit reliance on a single study of appropriate design to support its use as monotherapy, or as part of a different combination, for the same use. For example, a single study of a new combination vaccine designed to demonstrate adequate immune response will ordinarily provide sufficient evidence of effectiveness if the new combination contains products or antigens already proven to be effective alone or in other combinations. These situations are common for oncologic and antihypertensive drugs, but occur elsewhere as well.

e. Studies in a closely related disease

Studies in etiologically or pathophysiologically related conditions, or studies of a symptom common to several diseases (e.g., pain) can support each other, allowing initial approval of several uses or allowing additional claims based on a single adequate and well-controlled study. For example, certain anti-coagulant or anti-platelet therapies could be approved for use in two different settings based on individual studies in unstable angina/acute coronary syndrome and in the postangioplasty state. Because the endpoints studied and the theoretical basis for use of an anti-coagulant or anti-platelet drug are similar, each study supports the other for each claim. Similarly, single analgesic studies in several painful conditions would ordinarily be sufficient to support either a general analgesic indication or multiple specific indications. The recent approval of lamotrigine for treatment of Lennox-Gastaut Syndrome (a rare, largely pediatric, generalized seizure disorder) was based on a single adequate and well-controlled trial, due in part to related data showing efficacy of the drug in partial-onset seizures in adults.

f. Studies in less closely related diseases, but where the general purpose of therapy is similar

Certain classes of drug therapy, such as antimicrobials and antineoplastics, are appropriate interventions across a range of different diseases. For therapies of this type, evidence of effectiveness in one disease could provide independent substantiation of effectiveness in a quite different disease. For example, it is possible to argue that evidence of effectiveness of an antimicrobial in one infectious disease setting may support reliance on a single study showing effectiveness in other settings where the causative pathogens, characteristics of the site of infection that affect the disease process (e.g., structure and immunology) and patient population are similar.[6] Similarly, for an oncologic drug, evidence of effectiveness in one or more tumor types may support reliance on a single study showing effectiveness against a different kind of tumor, especially if the tumor types have a common biological origin.

g. Studies of different clinical endpoints

Demonstration of a beneficial effect in different studies on two different clinically meaningful endpoints could cross-substantiate a claim for

[6] See Division of Anti-Infective Drug Products: Points to Consider in the Clinical Development and Labeling of Anti-Infective Drug Products, October 1992.

effectiveness for each outcome. For example, the initial claim for effectiveness of enalapril for heart failure was supported by one study showing symptom improvement over several months and a second study showing improved survival in a more severely ill population. The two different findings, each from an adequate and well-controlled study, led to the conclusion that enalapril was effective in both treating symptoms and improving survival.

h. Pharmacologic/pathophysiologic endpoints

When the pathophysiology of a disease and the mechanism of action of a therapy are very well understood, it may be possible to link specific pharmacologic effects to a strong likelihood of clinical effectiveness. A pharmacologic effect that is accepted as a validated surrogate endpoint can support ordinary approval (e.g., blood pressure effects, cholesterol-lowering effects) and a pharmacologic effect that is considered reasonably likely to predict clinical benefit can support accelerated approval under the conditions described in 21 CFR 314 Subpart H and 21 CFR 601 Subpart E (e.g., CD4 count and viral load effects to support effectiveness of anti-viral drugs for HIV infection). When the pharmacologic effect is not considered an acceptable effectiveness endpoint, but the linkage between it and the clinical outcome is strong, not merely on theoretical grounds but based on prior therapeutic experience or well-understood pathophysiology, a single adequate and well-controlled study showing clinical efficacy can sometimes be substantiated by persuasive data from a well-controlled study or studies showing the related pharmacologic effect.

For example, a single clearly positive trial can be sufficient to support approval of a replacement therapy such as a coagulation factor, when it is combined with clear evidence that the condition being treated is caused by a deficiency of that factor. Demonstration of physical replacement of the deficient factor or restoration of the missing physiologic activity provides strong substantiation of the clinical effect. The corrective treatment of an inborn error of metabolism could be viewed similarly. In the case of preventive vaccines, one adequate and well-controlled clinical trial may be supported by compelling animal challenge/protection models, human serological data, passive antibody data, or pathogenesis information. The more evidence there is linking effects on the pharmacologic endpoint to improvement or prevention of the disease, the more persuasive the argument for reliance on a single clinical efficacy study.

Note, however, that plausible beneficial pharmacologic effects have often not correlated with clinical benefit, and, therefore, caution must be observed in relying on a pharmacologic effect as contributing to evidence

of effectiveness. For example, pharmacologic effects such as arrhythmia suppression by Type 1 antiarrhythmics and increased cardiac output by phosphodiesterase inhibitors or beta adrenergic inotropes resulted in increased mortality, rather than, as was expected, decreased sudden death and improved outcome in heart failure. The reasons for the absence of an expected correlation between pharmacologic and clinical effects are diverse and can include an incompletely understood relationship between the pharmacologic effect and the clinical benefit and the presence of other pharmacologic effects attributable to a drug in addition to the effect being measured and thought to be beneficial. Generally, the utility of pharmacologic outcomes in providing independent substantiation will be greatest where there is prior experience with the pharmacologic class. Even in this case, however, it is difficult to be certain that a pharmacologic effect that correlates with a clinical benefit accounts for all the clinical benefit or that other effects are not present and relevant.

3. Evidence of Effectiveness from a Single Study

When the effectiveness requirement was originally implemented in 1962, the prevailing efficacy study model was a single institution, single investigator, relatively small trial with relatively loose blinding procedures, and little attention to prospective study design and identification of outcomes and analyses. At present, major clinical efficacy studies are typically multicentered, with clear, prospectively determined clinical and statistical analytic criteria. These studies are less vulnerable to certain biases, are often more generalizable, may achieve very convincing statistical results, and can often be evaluated for internal consistency across subgroups, centers, and multiple endpoints.

The added rigor and size of contemporary clinical trials have made it possible to rely, in certain circumstances, on a single adequate and well-controlled study, without independent substantiation from another controlled trial, as a sufficient scientific and legal basis for approval. For example, the approval of timolol for reduction of post-infarction mortality was based on a single, particularly persuasive (low p-value), internally consistent, multicenter study that demonstrated a major effect on mortality and reinfarction rate. For ethical reasons, the study was considered unrepeatable. The Center for Biologics Evaluation and Research has also approved a number of products based upon a single persuasive study. The Agency provided a general statement in 1995 describing when a single, multicenter study may suffice (60 FR 39181; August 1, 1995), but the Agency has not comprehensively described the situations in which a single adequate and well-controlled study might be considered adequate support for an effectiveness claim, or the characteristics of a single study that could make it adequate support for an effectiveness claim.

Whether to rely on a single adequate and well-controlled study is inevitably a matter of judgment. A conclusion based on two persuasive studies will always be more secure than a conclusion based on a single, comparably persuasive study. For this reason, reliance on only a single study will generally be limited to situations in which a trial has demonstrated a clinically meaningful effect on mortality, irreversible morbidity, or prevention of a disease with potentially serious outcome and confirmation of the result in a second trial would be practically or ethically impossible. For example, sequential repetition of strongly positive trials that demonstrated a decrease in post-infarction mortality, prevention of osteoporotic fractures, or prevention of pertussis would present significant ethical concerns. Repetition of positive trials showing only symptomatic benefit would generally not present the same ethical concerns.

The discussion that follows identifies the characteristics of a single adequate and well-controlled study that could make the study adequate support for an effectiveness claim. Although no one of these characteristics is necessarily determinative, the presence of one or more in a study can contribute to a conclusion that the study would be adequate to support an effectiveness claim.

a. Large multicenter study

In a large multicenter study in which (1) no single study site provided an unusually large fraction of the patients and (2) no single investigator or site was disproportionately responsible for the favorable effect seen, the study's internal consistency lessens concerns about lack of generalizability of the finding or an inexplicable result attributable only to the practice of a single investigator. If analysis shows that a single site is largely responsible for the effect, the credibility of a multicenter study is diminished.

b. Consistency across study subsets

Frequently, large trials have relatively broad entry criteria and the study populations may be diverse with regard to important covariates such as concomitant or prior therapy, disease stage, age, gender or race. Analysis of the results of such trials for consistency across key patient subsets addresses concerns about generalizability of findings to various populations in a manner that may not be possible with smaller trials or trials with more narrow entry criteria. For example, the timolol postinfarction study randomized patients separately within three severity strata. The study showed positive effects on survival in each stratum supporting a conclusion that the drug's utility was not limited to a particular disease stage (e.g., relatively low or high severity).

c. Multiple *studies* in a single study

Properly designed factorial studies may be analyzed as a series of pairwise comparisons, representing, within a single study, separate demonstrations of activity of a drug as monotherapy and in combination with another drug. This model was successfully used in ISIS II, which showed that for patients with a myocardial infarction both aspirin and streptokinase had favorable effects on survival when used alone and when combined (aspirin alone and streptokinase alone were each superior to placebo; aspirin and streptokinase in combination were superior to aspirin alone and to streptokinase alone). This represented two separate (but not completely independent) demonstrations of the effectiveness of aspirin and streptokinase.

d. Multiple endpoints involving different events

In some cases, a single study will include several important, prospectively identified primary or secondary endpoints, each of which represents a beneficial, but different, effect. Where a study shows statistically persuasive evidence of an effect on more than one of such endpoints, the internal weight of evidence of the study is enhanced. For example, the approval of beta-interferon (Betaseron) for prevention of exacerbations in multiple sclerosis was based on a single multicenter study, at least partly because there were both a decreased rate of exacerbations and a decrease in MRI-demonstrated disease activity — two entirely different, but logically related, endpoints.

Similarly, favorable effects on both death and nonfatal myocardial infarctions in a lipid-lowering, postangioplasty, or postinfarction study would, in effect, represent different, but consistent, demonstrations of effectiveness, greatly reducing the possibility that a finding of reduced mortality was a chance occurrence. For example, approval of abciximab as adjunctive treatment for patients undergoing complicated angioplasty or atherectomy was supported by a single study with a strong overall result on the combined endpoint (decreased the combined total of deaths, new infarctions, and need for urgent interventions) and statistically significant effects in separate evaluations of two components of the combined endpoint (decreased new infarctions and decreased need for urgent interventions). In contrast, a beneficial effect on multiple endpoints that evaluate essentially the same phenomenon and correlate strongly, such as mood change on two different depression scales or SGOT and CPK levels postinfarction, does not significantly enhance the internal weight of the evidence from a single trial.

Although two consistent findings within a single study usually provide reassurance that a positive treatment effect is not due to chance, they do not protect against bias in study conduct or biased analyses. For example, a treatment assignment not well balanced for important prognostic variables could lead to an apparent effect on both endpoints. Thus, close scrutiny of study design and conduct are critical to evaluating this type of study.

e. Statistically very persuasive finding

In a multicenter study, a very low p-value indicates that the result is highly inconsistent with the null hypothesis of no treatment effect. In some studies it is possible to detect nominally statistically significant results in data from several centers, but, even where that is not possible, an overall extreme result and significance level means that most study centers had similar findings. For example, the thrombolysis trials of streptokinase (ISIS II, GISSI) had very sizable treatment effects and very low p-values, greatly adding to their persuasiveness. Preventive vaccines for infectious disease indications with a high efficacy rate (e.g., point estimate of efficacy of 80% or higher and a reasonably narrow 95% confidence interval) have been approved based on a single adequate and well-controlled trial.

4. Reliance on a Single, Multicenter Study — Caveats

While acknowledging the persuasiveness of a single, internally consistent, strong multicenter study, it must be appreciated that even a strong result can represent an isolated or biased result, especially if that study is the only study suggesting efficacy among similar studies. Recently, the apparent highly favorable effect of vesnarinone, an inotropic agent, in heart failure (60% reduction of mortality in what appeared to be a well-designed, placebo-controlled, multicenter trial with an extreme p-value) has proven to be unrepeatable. In an attempt to substantiate the finding, the same dose of the drug that seemed lifesaving in the earlier study significantly increased mortality (by 26%), and a lower dose also appeared to have a detrimental effect on survival. Although the population in the second study was, on the whole, a sicker population than in the first, the outcomes in similarly sick patients in each study were inconsistent so this factor does not explain the contradictory results.

When considering whether to rely on a single multicenter trial, it is critical that the possibility of an incorrect outcome be considered and that all the available data be examined for their potential to either support or undercut reliance on a single multicenter trial. In the case of vesnarinone, there were other data that were not consistent with the dramatically favorable outcome in the multicenter study. These data seemed to show an inverse dose-response relationship, showed no suggestion

of symptomatic benefit, and showed no effect on hemodynamic endpoints. These inconsistencies led the Agency, with the advice of its Cardio-Renal Advisory Committee, to refuse approval — a decision borne out by the results of the subsequent study.

This example illustrates how inadequacies and inconsistencies in the data, such as lack of pharmacologic rationale and lack of expected other effects accompanying a critical outcome, can weaken the persuasiveness of a single trial. Although an unexplained failure to substantiate the results of a favorable study in a second controlled trial is not proof that the favorable study was in error — studies of effective agents can fail to show efficacy for a variety of reasons — it is often reason not to rely on the single favorable study.

III. DOCUMENTATION OF THE QUALITY OF EVIDENCE SUPPORTING AN EFFECTIVENESS CLAIM

When submitting the requisite quantity of data to support approval of a new product or new use of an approved product, sponsors must also document that the studies were adequately designed and conducted. Essential characteristics of adequate and well-controlled trials are described in 21 CFR 314.126. To demonstrate that a trial supporting an effectiveness claim is adequate and well-controlled, extensive documentation of trial planning, protocols, conduct, and data handling is usually submitted to the Agency, and detailed patient records are made available at the clinical sites.

From a scientific standpoint, however, it is recognized that the extent of documentation necessary depends on the particular study, the types of data involved, and the other evidence available to support the claim. Therefore, the Agency is able to accept different levels of documentation of data quality, as long as the adequacy of the scientific evidence can be assured. This section discusses the factors that influence the extent of documentation needed, with particular emphasis on studies evaluating new uses of approved drugs.

For the purposes of this section, the phrase *documentation of the quality of evidence* refers to (1) the completeness of the documentation and (2) the ability to access the primary study data and the original study-related records (e.g., subjects' medical records, drug accountability records) for the purposes of verifying the data submitted as evidence. These interrelated elements bear on a determination of whether a study is adequate and well-controlled.

In practice, to achieve a high level of documentation, studies supporting claims are ordinarily conducted in accordance with good clinical practices (GCPs). Sponsors routinely monitor all clinical sites, and FDA routinely has access to the original clinical protocols, primary data, clinical site source documents for on-site audits, and complete study reports.

However, situations often arise in which studies that evaluate the efficacy of a drug product lack the full documentation described above (for example, full patient records may not be available) or in which the study was conducted with less monitoring than is ordinarily seen in commercially sponsored trials. Such situations are more common for supplemental indications because postapproval studies are more likely to be conducted by parties other than the drug sponsor and those parties may employ less extensive monitoring and data-gathering procedures than a sponsor. Under certain circumstances, it is possible for sponsors to rely on such studies to support effectiveness claims, despite less than usual documentation or monitoring. Some of those circumstances are described below.

A. Reliance on Less Than Usual Access to Clinical Data or Detailed Study Reports

FDA's access to primary data has proven to be important in many regulatory decisions. There are also reasons to be skeptical of the conclusions of published reports of studies. Experience has shown that such study reports do not always contain a complete, or entirely accurate, representation of study plans, conduct and outcomes. Outright fraud (i.e., deliberate deception) is unusual. However, incompleteness, lack of clarity, unmentioned deviation from prospectively planned analyses, or an inadequate description of how critical endpoint judgments or assessments were made are common flaws. Typically, journal article peer reviewers only have access to a limited data set and analyses, do not see the original protocol and amendments, may not know what happened to study subjects that investigators determined to be non-evaluable, and thus may lack sufficient information to detect critical omissions and problems. The utility of peer review can also be affected by variability in the relevant experience and expertise of peer reviewers. FDA's experiences with the Anturane Reinfarction Trial, as well as literature reports of the efficacy of tacrine and the anti-sepsis HA-1A antibody, illustrate its concerns with reliance on the published medical literature.

Notwithstanding these concerns, the presence of some of the factors discussed below can make it possible for FDA to rely on studies for which it has less than usual access to data or detailed study reports to partially or entirely (the so-called *paper* filing) support an effectiveness claim. FDA's reliance on a literature report to support an effectiveness claim is more likely if FDA can obtain additional critical study details. Section 1 below describes additional information that, if available, would increase the likelihood that a study could be relied on to support an effectiveness claim. Section 2 describes factors that may make efficacy findings sufficiently persuasive to permit reliance on the published literature alone. Note that the factors outlined in Section 2 are relevant to an assessment of the reliability of literature reports generally, whether alone, or accompanied by other important information as discussed in Section 1.

1. Submission of Published Literature or Other Reports in Conjunction with Other Important Information that Enhances the Reliability of the Data

If a sponsor wishes to rely on a study conducted by another party and cannot obtain the primary data from the study, for most well-conducted studies it is possible to obtain other important information, such as a protocol documenting the prospective plans for the trial, records of trial conduct and procedures, patient data listings for important variables, and documentation of the statistical analysis. FDA has considerable experience evaluating large multicenter outcome studies sponsored by U.S. and European government agencies (NIH, British Medical Research Council) and private organizations (the ISIS studies, the SAVE study) for which there was limited access to primary study data, but for which other critical information was available. Providing as many as possible of the following important pieces of information about a study, in conjunction with the published report, can increase the likelihood that the study can be relied on to support an effectiveness claim:

a. The protocol used for the study, as well as any important protocol amendments that were implemented during the study and their relation to study accrual or randomization.

b. The prospective statistical analysis plan and any changes from the original plan that occurred during or after the study, with particular note of which analyses were performed pre- and post-unblinding.

c. Randomization codes and documented study entry dates for the subjects.

d. Full accounting of all study subjects, including identification of any subjects with on-treatment data who have been omitted from analysis and the reasons for omissions, and an analysis of results using all subjects with on-study data.

e. Electronic or paper record of each subject's data for critical variables and pertinent baseline characteristics. Where individual subject responses are a critical variable (e.g., objective responses in cancer patients, clinical cures and microbial eradications in infectious disease patients, death from a particular cause), detailed bases for the assessment, such as the case report, hospital records, and narratives, should be provided when possible.

f. Where safety is a major issue, complete information for all deaths and drop-outs due to toxicity. For postapproval supplemental uses, however, there is generally less need for the results of lab tests or for details of adverse event reports and, consequently, much more limited documentation may be sufficient (e.g., only for unexpected deaths and previously undescribed serious adverse effects). Exceptions to this

approach would include situations in which the population for the supplemental use is so different that existing safety information has limited application (e.g., thrombolysis in stroke patients versus myocardial infarction patients) or where the new population presents serious safety concerns (e.g., extension of a preventive vaccine indication from young children to infants).

2. Submission of Published Literature Reports Alone

The following factors increase the possibility of reliance on published reports alone to support approval of a new product or new use:

a. Multiple studies conducted by different investigators where each of the studies clearly has an adequate design and where the findings across studies are consistent.

b. A high level of detail in the published reports, including clear and adequate descriptions of statistical plans, analytic methods (prospectively determined), and study endpoints, and a full accounting of all enrolled patients.

c. Clearly appropriate endpoints that can be objectively assessed and are not dependent on investigator judgment (e.g., overall mortality, blood pressure, or microbial eradication). Such endpoints are more readily interpreted than more subjective endpoints such as cause-specific mortality or relief of symptoms.

d. Robust results achieved by protocol-specified analyses that yield a consistent conclusion of efficacy and do not require selected post hoc analyses such as covariate adjustment, subsetting, or reduced data sets (e.g., analysis of only responders or compliant patients, or of an "eligible" or “evaluable” subset).

e. Conduct of studies by groups with properly documented operating procedures and a history of implementing such procedures effectively.

There have been approvals based primarily or exclusively on published reports. Examples include the initial approval of secretin for evaluation of pancreatic function and recent approvals of bleomycin and talc for malignant pleural effusion and doxycycline for malaria.

B. Reliance on Studies with Alternative, Less Intensive Quality Control/On-Site Monitoring

Industry-sponsored studies typically use extensive on-site and central monitoring and auditing procedures to assure data quality. Studies supported by other sponsors may employ less stringent procedures and may use no on-site monitoring at all. An International Conference on Harmonisation guideline on good clinical practices,[7] recently accepted internationally, emphasizes that the extent of monitoring in a trial should be based on trial-specific factors (e.g., design, complexity, size, and type of study outcome measures) and that different degrees of on-site monitoring can be appropriate. In recent years, many credible and valuable studies conducted by government or independent study groups, often with important mortality outcomes, had very little on-site monitoring. These studies have addressed quality control in other ways, such as by close control and review of documentation and extensive guidance and planning efforts with investigators. There is a long history of reliance on such studies for initial approval of drugs as well as for additional indications. Factors that influence whether studies with limited or no monitoring may be relied on include the following:

1. The existence of a prospective plan to assure data quality.

2. Studies that have features that make them inherently less susceptible to bias, such as those with relatively simple procedures, noncritical entry criteria, and readily assessed outcomes.

3. The ability to sample critical data and make comparisons to supporting records (e.g., hospital records).

4. Conduct of the study by a group with established operating procedures and a history of implementing such procedures effectively.

[7] International Conference on Harmonisation Guidance for Industry E6, *Good Clinical Practice: Consolidated Guideline*, April 1996.